Structure of Human Osteonectin Based upon Analysis of cDNA and Genomic Sequences[†,‡]

Xavier C. Villarreal, Kenneth G. Mann, and George L. Long*

Department of Biochemistry, University of Vermont, Burlington, Vermont 05405

Received November 29, 1988; Revised Manuscript Received April 13, 1989

ABSTRACT: Overlapping human bone osteonectin cDNAs were obtained by screening two independent human SaOS-2 λgt11 libraries using antibovine osteonectin monoclonal antibodies. One clone contains a 0.54-kb insert and the other a 1.9-kb insert. Insertion fragments from λ clones were liberated by restriction digestion and subcloned into pUC19 for sequencing. Digestion of the 1.9-kb insert with *Eco*RI released 0.4- and 1.5-kb fragments. Sequencing analysis revealed that the 0.54- and 0.4-kb fragments are identical except for 150 nucleotides missing at the 5′ region of the 0.4-kb fragment. The composite nucleotide sequence of human osteonectin has a total length of 2091 nucleotides and is comprised of 50 nucleotides of 5′-noncoding sequence, a coding segment for 303 amino acids, a termination codon, and 1114 nucleotides of 3′-noncoding sequence. The primary transcript codes for 286 amino acids of mature protein and a 17-residue amino-terminal hydrophobic signal peptide. Outstanding properties inferred from the primary structure are putative Ca^{2+} binding domains located in the glutamic acid rich NH_2 terminus (residues 1–52) and two "EF"-hand domains in the C-terminal half of the protein (residues 165–176 and 257–286). The mature protein also contains a cysteine-rich, highly hydrophilic region homologous to the ovomucoid serine protease inhibitors (residues 76–132). Overlapping human genomic clones in λEMBL3 for osteonectin have been isolated and characterized. Intron/exon junction sequencing of the human osteonectin gene shows the presence of 10 exons and 9 introns. The mature protein is encoded by nine exons separated by eight introns. Exonic sequences are in complete agreement with that of cDNA from the SaOS-2 cell line. The position of introns is in good agreement with predicted protein domains. Comparison at both the nucleic acid and translated amino acid level demonstrates human bone osteonectin to be identical with human placental SPARC and very similar to bovine osteonectin and mouse SPARC. These findings clearly establish that the nascent translation product for bone osteonectin and SPARC from nonmineralized tissue are identical.

Osteonectin is a 32000 molecular weight single-chain acidic glycoprotein, originally purified as a noncollagenous component from EDTA-solubilized extracts from fetal and adult bovine bone matrix (Termine et al., 1981; Romberg et al., 1985). Studies performed in vitro using either fetal or adult bovine osteonectin suggest that bone mineralization may be regulated by osteonectin, because osteonectin has high affinity for Ca^{2+}, hydroxyapatite, and type I collagen (Termine et al., 1981; Romberg et al., 1985, 1986).

Expression experiments using polyclonal antibodies against fetal bone osteonectin (Wasi et al., 1984) or cDNA probes (Young et al., 1986) have suggested the presence of osteonectin in a variety of nonmineralized tissues and cultured cells. Stenner and co-workers have reported the presence of osteonectin in human platelets (Stenner et al., 1986). These results have led to a reevaluation of the biological role for osteonectin. Recently, several homologous noncollagenous proteins have been described. A recent review of the osteonectin family of proteins has appeared (Tracy et al., 1988). These include 43K protein, secreted by bovine aortic endothelial cells when subjected to cellular stress (Sage et al., 1986), BM-40, a mouse basement membrane protein extracted from the Englebreth–Holm–Swarm tumor (Mann et al., 1987), and SPARC (secreted protein, acidic and rich in cysteine), derived from the mouse embryo parietal endoderm (Mason et al., 1986) and human placenta (Swaroop et al., 1988). Observations obtained from purification, expression, and distribution of osteonectin and osteonectin homologues in tissue matrices suggest that osteonectin may participate in a general manner in the formation and regulation of extracellular matrix.

In this study, two cDNA clones encoding human osteonectin from SaOS-2 λgt11 libraries and two human genomic clones from an λEMBL library were obtained. Analyses of both the genome for intron/exon structure and the cDNA for salient structural/functional regions in the translated polypeptide are described and compared to other members of the protein family.

Materials and Methods

λgt11 Expression. λgt11 SaOS-2 libraries were kindly supplied by Drs. Mark A. Thiede, Merck, Sharp & Dohme Research Laboratories, West Point, PA, and Mitchell J. Weiss, Department of Human Genetics, University of Pennsylvania. Phage were grown on agar plates and screened by the method of Young and Davis (1983), using a biotin–avidin–peroxidase amplification protocol outlined by French et al. (1986). Plating phage densities for initial and subsequent screens were approximately 150000 PFUs/150-mm petri dish, 3000 PFUs/150-mm dish, and 20–100 PFUs/100-mm dish.

Triton-free nitrocellulose filters were rinsed in Tris-buffered saline–Tween 20 (TBST: 50 mM Tris, pH 7.9, 150 mM NaCl, and 0.05% Tween 20) and blocked (2% ovalbumin and 0.05% NaN_3 in TBST) prior to exposure to the primary antibody solution containing 2 μg/mL antibovine osteonectin IgG

[†]Partially supported by USPHS Grants P01 AG04875, R01 HL38899, and C06 HL39745 from the National Institutes of Aging and Heart, Lung and Blood.

[‡]The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02863.

* To whom correspondence should be addressed.

0006-2960/89/0428-6483$01.50/0 © 1989 American Chemical Society

monoclonal antibody (Stenner et al., 1984). Filters were subsequently treated with biotinylated goat antimouse IgG secondary antibody and the avidin–peroxidase reporter system according to the manufacturer (Clontech Laboratories, Palo Alto, CA).

λgt11 Purification and Subcloning. Isolated phage DNA was prepared by the plate lysis–CsCl banding–formamide method outlined elsewhere (Davis et al., 1980). The DNA following ethanol precipitation was suspended in TE (10 mM Tris, pH 7.5/1 mM Na_2EDTA) and stored at 4 °C.

Phage DNA was digested with appropriate restriction enzymes and electrophoresed in a 0.6% low-melt-temperature agarose (Bio-Rad Laboratories, Richmond, CA) in TAE buffer (0.04 M Tris–acetate/0.001 M Na_2EDTA, pH 8.0) and stained with ethidium bromide. Desired fragments were isolated from agarose by the method of Ogden and Adams (1987), ligated into cleaved plasmid pUC19, and used to transform competent DH5α cells (Bethesda Research Laboratories) according to the supplier. Plasmid from positive colonies was mini-prepped by the method of Birnboim and Doly (1979) and analyzed by standard procedures (Maniatis et al., 1982).

DNA Sequencing. Cesium chloride purified plasmid prepared according to Katz et al. (1973) or mini-prepped DNA prepared by a modification of Birnboim and Doly (1979) was used for sequencing. Restriction fragments for Maxam–Gilbert reactions were obtained by electrophoresing on a 3.5% polyacrylamide gel and crush elution and ethanol precipitation (Maxam & Gilbert, 1980). Overhanging 5′ ends were treated with bacterial alkaline phosphatase and kinased by T4 polynucleotide kinase as suggested by the supplier (Bethesda Research Laboratories). Labeled end fragments were cleaved, resolved on a 3.5% polyacrylamide gel, and purified by crush elution. Sanger dideoxy sequencing was performed by using T7 DNA polymerase (Sequenase, from United States Biochemical Corp., Cleveland, OH) with the alterations described by Zhang et al. (1988). Six percent polyacrylamide/7 M urea sequencing gels were dried down (Sanger & Coulson, 1977), and 20% gels were fixed in 5% acetic acid prior to being exposed to Kodak X-omat AR X-ray film with the aid of a DuPont Cronex Plus intensifying screen. All DNA sequences were analyzed with the aid of the University of Wisconsin Genetics Computer Group sequence analysis software package, version 5.

Oligonucleotide Synthesis. Oligonucleotides used as primers for Sanger sequencing were synthesized by using 2-cyanoethyl phosphoramidite derivatives on an Applied Biosystems Inc. (Foster City, CA) DNA Model 381A synthesizer in the trityl-off mode. Oligonucleotides released from the synthesis columns were either gel purified from a 20% sequencing gel or used directly following lyophilization.

Cloning of Genomic Human Osteonectin DNA. A library of partially *Sau*3A digested human genomic DNA in bacteriophage λEMBL3 was kindly provided by Dr. R. T. A. MacGillivray, Department of Biochemistry, University of British Columbia, Canada, and is described elsewhere (Geddes et al., 1989). The library was initially screened by a modification of the procedure of Benton and Davis (1977) at a density of 5×10^4 plaques per 150-mm petri dish (12 dishes). Radiolabeled probe was prepared by the random priming method of Feinberg and Vogelstein (1983) using $dCT^{32}P$ and the 0.54-kb *Eco*RI human osteonectin cDNA fragment from pHVON-9-2 described below. Sequential screenings were performed on all positive clones at increasingly lower plaque densities until isolated positive plaques were obtained. Genomic fragments were subcloned into plasmid pUC19 and analyzed essentially as described above.

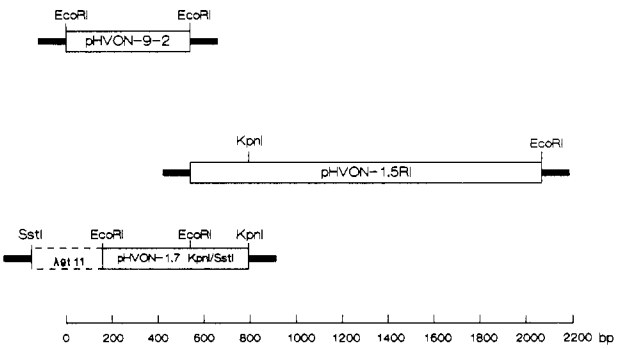

FIGURE 1: Diagrammatic illustration of human osteonectin cDNA subclones in pUC19 plasmid. Solid bar represents pUC19 vector; dashed box, a portion of the λgt11 vector; open box, human osteonectin cDNA insert. Only human osteonectin cDNAs are drawn to scale. Nucleotide numbering is based upon the sequence shown in Figure 2.

RESULTS

λgt11 cDNA Clones for Human Osteonectin. Initial screening of 1.5×10^6 λgt11 phage (10 137-mm plates) from a SaOS-2 human osteosarcoma library resulted in 3 clones each containing a 0.54-kb insert (Figure 1). The insert was liberated by *Eco*RI digestion and subcloned into the *Eco*RI restriction site of pUC19 and sequenced. As no full-length cDNA was obtained, a second SaOS-2 λgt11 library was screened. From 6×10^6 λgt11 phage (21 137-mm plates), 1 clone possessing a 1.9-kb insert was isolated. Fragments of 0.4 and 1.5 kb were released by *Eco*RI digestion of the 1.9-kb insert and subcloned into pUC19 (Figure 1). To establish the orientation of the 0.4- and 1.5-kb fragments, a *Kpn*I/*Sst*I overlapping fragment was also subcloned into pUC19 plasmid (Figure 1).

Analysis of cDNA Sequencing. Nucleotide sequencing revealed that the 0.4-kb fragment from the 1.9-kb cDNA clone is identical with that of the 0.54-kb clone except that it lacks 159 nucleotides at the 5′ end. The 1.5-kb fragment corresponds to a region 3′ and adjacent to the 0.54-kb fragment. The 2091-nucleotide composite sequence for human osteonectin cDNA (and *Eco*RI linkers) and the translated protein product are displayed in Figure 2. The complete human osteonectin cDNA consists of 50 nucleotides of 5′-noncoding sequence, 909 nucleotides coding for a 17 amino acid signal peptide and 286 amino acids of the mature protein, a single TAA termination codon, and 1114 nucleotides of 3′-noncoding sequence. A poly(A) recognition signal (AATAAA) exists at position 2077 (see Figure 2). The nucleotide composition of human ostenectin consists of 24.6% G, 24.3% A, 25.5% T, and 25.6% C.

Figure 3 compares the nucleotide sequence of human osteonectin cDNA with those of bovine osteonectin and mouse SPARC. The comparison clearly establishes a homologous relationship between these proteins and their genetic material.

Structure of Human Osteonectin. For primary and secondary structural analysis, a composite diagram of the secondary structure of human osteonectin emphasizing charged and cysteinyl residues and Chou–Fasman (1978) predictions is shown in Figure 4. On the basis of the NH_2-terminal sequence of human bone osteonectin reported by Fisher et al. (1987a), the translated cDNA may be deduced to code for a 17 amino acid leader peptide sequence followed by 286 amino acids of mature protein. The derived amino acid composition of mature human osteonectin is identical with that reported

```
                                                                    ▼A
1   GAA TTC CGC GCG CTC TGC CTG CCG CCT GCC TGC CTG CCA CTG AGG GTT CCC AGC
      -17                           -10                 -5
55  AGC ATG AGG GCC TGG ATC TTC TTT CTC CTT TGC CTG GCC GGG AGG GCC TTG GCA
    Met Arg Ala Trp Ile Phe Phe Leu Leu Cys Leu Ala Gly Arg Ala Leu Ala
+1      ▼B      5                  10                  15
109 GGC CCT CAG CAA GAA GCC CTG CCT GAT GAG ACA GAG GTG GTG GAA GAA ACT GTG
    Ala Pro Gln Gln Glu Ala Leu Pro Asp Glu Thr Glu Val Val Glu Glu Thr Val
     20          ▼C     25                  30                  35
163 GCA GAG GTG ACT GAG GTA TCT GTG GGA GCT AAT CCT GTC CAG GTG GAA GTA GGA
    Ala Glu Val Thr Glu Val Ser Val Gly Ala Asn Pro Val Gln Val Glu Val Gly
                40                  45                  50          ▼D
217 GAA TTT GAT GAT GGT GCA GAG GAA ACC GAA GAG GAG GTG GTG GCG GAA AAT CCC
    Glu Phe Asp Asp Gly Ala Glu Glu Thr Glu Glu Glu Val Val Ala Glu Asn Pro
     55                  60                  65                  70
271 TGC CAG AAC CAC CAC TGC AAA CAC GGC AAG GTG TGC GAG CTG GAT GAG AAC AAC
    Cys Gln Asn His His Cys Lys His Gly Lys Val Cys Glu Leu Asp Glu Asn Asn
              75                  80                  85                  ◇ 90
325 ACC CCC ATG TGC GTG TGC CAG GAC CCC ACC AGC TGC CCA GCC CCC ATT GGC GAG
    Thr Pro Met Cys Val Cys Gln Asp Pro Thr Ser Cys Pro Ala Pro Ile Gly Glu
          ▼E    95                  100                 105
379 TTT GAG AAG GTG TGC AGC AAT GAC AAC AAG ACC TTC GAC TCT TCC TGC CAG TTC
    Phe Glu Lys Val Cys Ser Asn Asp Asn Lys Thr Phe Asp Ser Ser Cys His Phe
     110                 115         ◇       120                 125
433 TTT GCC ACA AAG TGC ACC CTG GAG GGC ACC AAG AAG GGC CAC AAG CTC CAC CTG
    Phe Ala Thr Lys Cys Thr Leu Glu Gly Thr Lys Lys Gly His Lys Leu His Leu
          130                 ▼F                  140
487 GAC TAC ATC GGG CCT TGC AAA TAC ATC CCC CCT TGC CTG GAC TCT GAG CTG ACC
    Asp Tyr Ile Gly Pro Cys Lys Tyr Ile Pro Pro Cys Leu Asp Ser Glu Leu Thr
    145                 150                 155                 160
541 GAA TTC CCC CTG CGC ATG CGG GAC TGG CTC AAG AAC GTC CTG GTC ACC CTG TAT
    Glu Phe Pro Leu Arg Met Arg Asp Trp Leu Lys Asn Val Leu Val Thr Leu Tyr
              165                 170                 175                 ▼G 180
595 GAG AGG GAT GAG GAC AAC AAC CTT CTG ACT GAG AAG CAG AAG CTG CGG GTG AAG
    Glu Arg Asp Glu Asp Asn Asn Leu Leu Thr Glu Lys Gln Lys Leu Arg Val Lys
                    185                 190                 195
649 AAG ATC CAT GAG AAT GAG AAG CTC CTG GAG GCA GGA GAC CAC CCC GTG GAG CTG
    Lys Ile His Glu Asn Glu Lys Leu Leu Glu Ala Gly Asp His Pro Val Glu Leu
     200                 205                 210                 215
703 CTG GCC CGG GAC TTC GAG AAG AAC TAT AAC ATG TAC ATG TTC CCT GTA CAC TGG
    Leu Ala Arg Asp Phe Glu Lys Asn Tyr Asn Met Tyr Ile Phe Pro Val His Trp
                220                 225         ▼H      230
757 CAG TTC GGC CAG CTG GAC CAG CAG CCC ATT GAC GGG TAC CTC TCC CAC ACC GAG
    Gln Phe Gly Gln Leu Asp Gln His Pro Ile Asp Gly Tyr Leu Ser His Thr Glu
    235                 240                 245                 250
811 CTG GCT GCA CTG CGT GCT CCC CTG ATC GCG ATG GAG CAT TGC ACC ACC CGC CTT
    Leu Ala Pro Leu Arg Ala Pro Leu Ile Pro Met Glu His Cys Thr Thr Arg Leu
              255                 260                 265                 270
865 TTC GAG ACC TGT GAC CTG GAC AAT GAC AAG TAC ATG GCC CTG GAT GAG TGG GCC
    Phe Glu Thr Cys Asp Leu Asp Asn Asp Lys Tyr Ile Ala Leu Asp Glu Trp Ala
                    275         ▼|      280                 285
919 GGC TGC TTC GCC ATC AAG CAG AAG GAT ATC GAC AAG GAT CTG GTG ATC TAA ATC
    Gly Cys Phe Gly Ile Lys Gln Lys Asp Ile Asp Lys Asp Leu Val Ile End

973  CAG TGC TTC CAC AGT ACG GGA TTC TCT CTT TAA CCC TCG CCT TCG TGT TTC CCC

1027 CAA TGT TTA AAA TGT TTG GAT GGT TTG TTG TTG TGC CTG GAG ACA AGG TGC TAA

1081 CAT AGA TTT AAG TGA ATA CAT TAA CGG TGC TAA AAA TGA AAA TTG TAA GCC AAG

1135 ACA TGA CAT TCT TAG CTG TAA CTT AAC TAT TAA GGC GTT TTC CAC ACG CAT TAA

1189 TAG TCC CAT TTT TCT CTT GCC ATT TGT AGG TTT GCC CAT TGT CTT ATT GGG ACA

1243 TGG GTG GAG ACG GAT CTG CTG GGC TCT GCT TAA ACA CAG ATT GCA GCT TCA ACT

1297 TTT CTC TTT AGT GTT CTG TTT GAA ACT AAT ACT TAG CGA GTC AGA CTT TGT GTT

1351 CAT TTC ATT TCA GGG TGT TGG CTG CGT GTG GGC TTC CGC AGG TGG CCT GGA GGT

1405 GGG GAA AGG GAA GTA ACA GAC AGA CGA TGT TGT CAA GGA TGG TTT TGG GAC TAG

1459 AGG CTC AGT GGT GGG AGA GAT CCC TGC AGA ACG CAC CAA CCA GAA CGT GGT TTG

1513 CCT GAG GCT GTA ACT GAG AGA AAG ATT CTG GGG CTG TGT TAT GAA AAT ATA GAC

1567 ATT CTC ACA TAA GGC CAG TTC ATC ACC ATT TGC TCC TTT ACC TTT CAG TGC AGT

1621 TTC TTT TCA CAT TAG GCT GTT GGT TCA AAC TTT TGG GAG CAC GGA CTG TCA GTT

1675 CTC TGG GAA GTG GTC AGC GCA TCC TGC AGG GCT TCT CCT CCT CTG TCT TTT GGA

1729 GAA CCA GGG GTG TTC TCA GGG GCT CTA GGG AGT GCC AGG CTG TTT CAG CCA GGA

1783 AGG CCA AAA TCA AGA GTG AGA TGT AGA AAG TTG TAA AAT AGA AAA AGT GGA GTT

1837 GGT GAA TCG GTT GTT CTT TCG TCA CAT TTG GAT GAT TGT CAT AAG GTT TTT AGC

1891 ATG TTC CTC CTT TTC TTC ACG CTC GCC CTT TTT CTT CTA TTA ATC AAG AGA AAC

1945 TTC AAA GTT AAT GGG ATG GTC GGA TCT CAC AGG CTG AGA ACT GCT TCA CCT CCA

1999 AGG ATT TCA TGA AAA AGC TGC TTC TTA TTA ATC ATA CAA ACT CTC ACC ATG ATG

2053 TGA AGA GTT TCA CAA ATC TTT CAA AAT AAA AGG AAT TC
```

FIGURE 2: Nucleotide and translated amino acid sequence of human osteonectin cDNA. Nucleotides 1–7 and 2084–2090 are parts of the *Eco*RI cloning site linkers and are not considered part of the cDNA. The signal peptide is shown in italics, bold downward arrows identify intron/exon junctions, bold capital letters refer to introns, open upward arrows identify potential N-glycosylation sites, asterisks denote potential O-phosphorylation sites, and the poly(A) recognition signal is underlined. Amino acid numbering is based upon the amino-terminal residue of the mature protein being number 1. Sequencing of human genomic DNA is in complete agreement with nucleotides 46–1258 shown above.

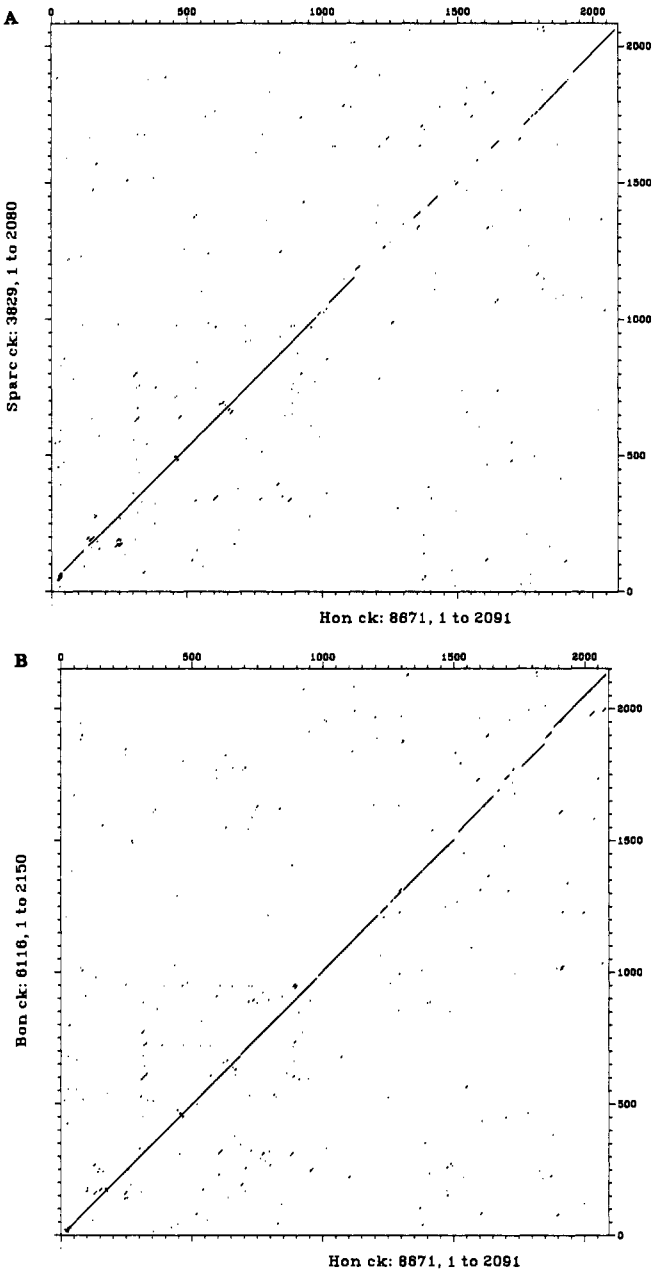

FIGURE 3: Graphical comparisons of human osteonectin with bovine osteonectin and mouse SPARC cDNAs. The nucleotide sequence for human osteonectin (Hon 1–2082) is aligned with that of mouse SPARC (1–2080) (panel A) and bovine osteonectin (Bon 1–2180) (panel B) using the COMPARE program of Maziel and Lenck (1981). Regions of sequence similarity (14 identical nucleotides in a segment of 21) are plotted as points.

for human placental SPARC (Swaroop et al., 1988) and consists of 13 Gly, 14 Ala, 20 Val, 25 Leu, 11 Ile, 12 Phe, 3 Trp, 7 Try, 17 Thr, 7 Ser, 8 Arg, 20 Lys, 12 His, 23 Asp, 14 Asn, 34 Glu, 10 Gln, 18 Pro, 4 Met, and 14 Cys.

The amino acid composition predicted for the human osteonectin molecule is in general agreement with that derived from chemical analysis (Fisher et al., 1987a). Notable differences between amino acid and cDNA sequence analyses are the values obtained for cysteine, methionine, and serine. The low yields calculated from amino acid analysis for cysteine and methionine may reflect losses due to oxidation. More difficult to explain is the higher serine content indicated by chemical analysis. The high estimate from amino acid composition may have arisen from contaminating buffer components or a small peptide.

Kyte and Doolittle (1982) analysis reveals that the signal peptide is typical in being strongly hydrophobic whereas the mature protein consists of several strongly hydrophilic regions

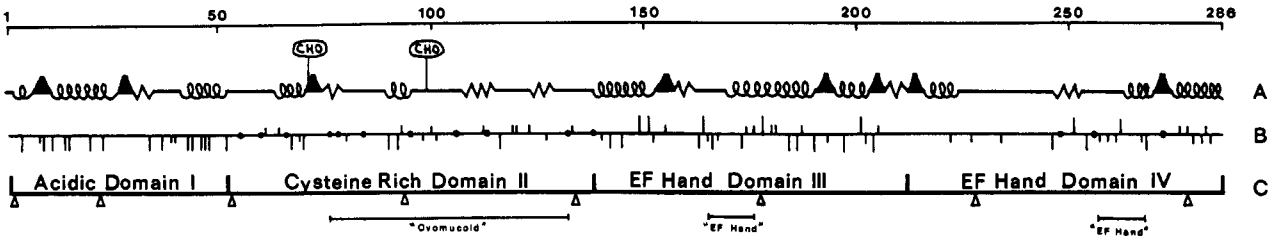

FIGURE 4: Graphical representation of human osteonectin. Residue numbering is for the mature protein and is the same as that in Figure 2. (A) Chou–Fasman prediction of secondary structure. Horizontal lines represent random-coil, β-turns are shown by closed triangles, α-helix by coils, and β-strand by sawteeth. Potential N-glycosylation sites are denoted by CHO. (B) Charge and cysteine distribution. Cysteine residues are shown with cloned circles. Positively charged Lys and Arg residues are represented by upward short and long vertical bars, and negatively charged Asp and Glu are represented by downward short and long bars, respectively. (C) Protein structural/functional domains. Positions of introns in the corresponding gene are shown by open triangles.

and lacks any hydrophobic segments. Chou–Fasman analysis predicts the mature protein to be 38% α-helical, 14% β-strand, and 48% random-coil or β-turn.

On the basis of charge density, cysteine distribution, and potential secondary structure analysis, the final processed polypeptide may be divided into four distinct structural/functional domains. Domain I, consisting of residues 1–52, is highly acidic and predicted to be predominantly α-helical. Two internal homologous repeats (residues 1–22 and 28–50 with sequence identity of 39%) are present (Figure 5). Domain I is dominated by glutamic acid, with 15 of the total 34 glutamates found in this 52-residue segment. This region lacks basic (lysine and arginine) and sulfhydryl (cysteine) amino acids. Domain I of the protein may be involved in hydroxyapatite interactions.

A second region comprised of residues 53–138 is characterized by the following features: (a) Secondary structural analysis suggests that it is mainly irregular in secondary structure and strongly hydrophilic. (b) It is cysteine rich, possessing 11 of the 14 cysteine residues found in the mature protein. (c) It has two potential Asn-X-Thr/Ser N-glycosylation sites, located at positions 71 and 99. The second domain has a high concentration of proline (8 of 18 total prolines) and serine (4 of 6 serines). The presence of phosphoserine in human osteonectin has been reported by Fisher et al. (1987). Domain II contains a potential serine phosphorylation site (residue 104). Domain II also contains a segment (residues 76–132, Figure 5) homologous to the ovomucoid serine protease inhibitors (Laskowski et al., 1987). The degree of identity with these inhibitors is approximately 30% and is comparable to that between other members of the ovomucoid family. The argument for homology is particularly strengthened by the position of cysteine residues in osteonectin relative to those in the protease inhibitors. At present, no protease inhibitor activity has been demonstrated for bovine osteonectin.

The third domain consists of amino acids 139–212. Chou–Fasman and Kyte–Doolittle predictions suggest that this area is largely α-helical and hydrophilic in structure. One prominent feature of the third domain is a 12 amino acid segment homologous to the calcium binding "EF-hand" found in the calmodulin family of proteins (Kretsinger, 1980). This potential calcium binding region is defined by residues 165–176 (see Figure 5). Secondary structural analysis suggests that α-helices are adjacent to the "EF-hand" which may aid as a stabilizing force in calcium binding (refer to Chou–Fasman predictions depicted in Figure 4). Domain III also contains at its amino terminus (residue 141) a second potential serine phosphorylation site.

The last domain includes amino acids 213–286. Secondary structural analysis predicts that the fourth domain is also composed of several α-helical regions. A potential second calcium binding "EF-hand" loop is positioned at residues 257–268 and is surrounded by potential α-helical regions.

Human Genomic Clones for Osteonectin. Screening of $\sim 3 \times 10^5$ independent human genomic/λEMBL3 plaques with a 540 bp cDNA probe from plasmid pHVON-9-2 resulted in 7 positive clones. Fragments from two of the clones, λ5 and λ8, were subcloned into plasmid pUC19 and further characterized by restriction enzyme mapping and DNA sequencing. As shown in Figure 6, the two clones overlap one another by $\sim$8 kb and span a total genomic DNA segment of $\sim$21 kb. Coding portions of the gene and all intron/exon junctions were sequenced by employing an "intron-jumping" technique that utilizes custom oligonucleotide primers. The intron-jumping strategy consisted of sequencing genomic DNA from within a region corresponding to cDNA sequence until the sequence extended into an intronic region. On the basis of this now-established intron/exon junction, a new set of overlapping sequencing primers were synthesized to the predicted adjacent exonic region and used to sequence from this point in both directions. To date, approximately 4 kb of human osteonectin genomic DNA has been sequenced. Clones λ5 and λ8 contain sequences identical with cDNA nucleotides 46–1258 (Figure 2), including the entire protein-coding portion of the cDNA. Segments of genomic DNA corresponding to the cDNA sequence are interrupted by nine introns, as shown in Figure 2. The first intron resides in the 5′-noncoding region of the cDNA, and all of the remaining introns are within the protein-coding portion of the cDNA. All of the intron/exon junctions have been sequenced from both directions except for intron A in the 5′-noncoding region. Genomic clone λ5 does not extend upstream of intron A; consequently, sequencing across the intron/exon junction from the upstream side is not possible with the clones presently available. On the basis of the location of cDNA nucleotide 46 in λ5 (see Figure 6), the 5′-noncoding intron A can be estimated to be greater than 6 kb in length, and the size of the entire gene can be estimated to be greater than 21 kb in length.

The intronic splice site recognition sequences and intronic types are shown in Table I for the nine introns. Table I shows that all of the introns contain the dinucleotides GT and AG at the 5′ and 3′ ends, respectively, in accordance with the pattern first noted by Breathnach et al. (1978). Adjacent nucleotide sequences closely resemble the consensus splice sites originally defined by Mount (1982). Segments closely upstream of the 3′ intron splice site also resemble the consensus sequence for the intron splicing branch site proposed by Sharp and co-workers (Sharp, 1987).

Translation of the inverse complement of the sequence in Figure 2 reveals a long open reading frame (281 amino acids) for nucleotides 8–849. However, analysis of the intron splice

FIGURE 5: Comparison of human osteonectin amino acid sequence to homologous proteins. Amino acids are shown with the one-letter code. Abbreviations: HON, human osteonectin; BON, bovine osteonectin; OPI, turkey ovomucoid protease inhibitor, domain III; consensus EF, "EF-hand"; HS, human skeletal troponin C; circumflex, a residual addition; dashes, amino acid deletion. For BON and mouse SPARC, only differences are shown relative to HON. For OPI and HS, the identities to human osteonectin are shown with bold letters. Similarities apparently do not extend beyond the noted regions.

junctions and flanking sequences shown in Table I reveals that it would not be possible to generate the proper mRNA for the inverse complement open reading frame.

Discussion

Osteonectin was first isolated and characterized by Termine et al. (1981a) from fetal calf bone, using a denaturing guanidine/EDTA extraction procedure. The resulting protein had an apparent molecular weight of 32 000, based upon gel filtration in 4 M guanidine hydrochloride, and contained ~5% carbohydrate and ~0.5% organic phosphorus. Analysis of amino acid composition showed it to be unusually rich in acidic

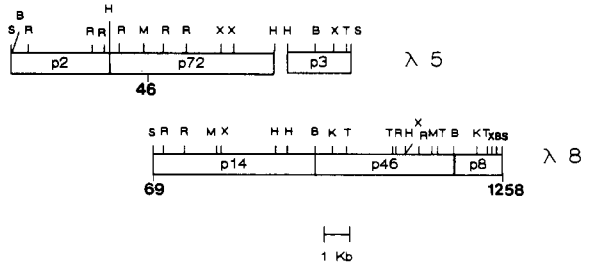

FIGURE 6: Restriction endonuclease map of human osteonectin genomic clones. Horizontal bars represent two overlapping human genomic DNA inserts in phage λEMBL3. Numbers within the bars refer to subclones in plasmid pUC19, with vertical broad lines noting the ends of each subclone. Numbers below the ends of clone λ8 and within clone λ5 refer to cDNA nucleotide numbers in Figure 2. Approximately 6 kb on the 5′ (left) end of clone λ5 is believed to be a part of the 5′-noncoding intron A shown in Figure 2. Abbreviations: S, *Sal*I; B, *Bam*HI; R, *Eco*RI; H, *Hin*dIII; M, *Sma*I; X, *Xba*I; T, *Sst*I; K, *Kpn*I. *Sal* and *Bam* sites at the ends of clones λ5 and 8 are parts of the multiple cloning sites of λEMBL3.

Table I: Summary of Human Osteonectin Introns

intron	sequence[a]	type[b]
	20-50	
Consensus	5′-AAG↓GTRAGT.....YNYRAY.....Y_nNCAG↓G- 3′	
	C	
	16	
A	.....CTCAA TC TCTC CTCAATAG GG	-
	34	
B	CCT GTAAGT.....CCTGAC..CTCTTTCCCACTAG CA	0
	43	
C	GAG GTATGT.....TGGGAT..TCCCTCAATTCCAG GT	0
	31	
D	AAA GTATGT.....GGTGAT..TTTCTGCCCTACAG AT	I
	30	
E	AAG GTGAGG.....GGGAAT....CTTTGCCAACAG GT	0
	24	
F	AAT GTGAGT.....GCTGAC....TTTGGAACCTAG AC	I
	56	
G	CGG GTAAGT.....CCCGAC...TCTACTCCCTCAG GT	0
	35	
H	CGG GTAAGA.....TTCCAT..CTCCTTCCCTGCAG GT	II
	25	
I	AGA GTGAGT.....TTTAAC...TCTTGCTTTGCAG AG	I

[a] Nucleotide sequence for each of the human osteonectin introns is aligned with that of the consensus sequence reported by Mount (1982). Splice sites are noted with vertical arrows, with the intron positioned between and the exons flanking the arrows. The intronic recognition segment approximately 20–50 nucleotides in the 5′ direction from the 3′ ends of introns and the adenine nucleotide (underlined) believed to be involved in forming the intermediate lariat structure of gene splicing (Sharp, 1987) are also shown. Numbers above each sequence denote the distance in nucleotides from the 3′-splice site to the putative recognition segment. Abbreviations: dots, additional intronic sequence not presented; R, purine; Y, pyrimidine; N, purine or pyrimidine. [b] Type intron based upon whether it occurs between (O), or after the first (I) or second (II) nucleotide of the corresponding cDNA triplet codon, according to Sharp (1981).

(Asp and Glu) residues. Osteonectin has more recently been isolated from bovine (Romberg et al., 1985) and human (Fisher et al., 1987a) bone tissue in the absence of denaturing agents. Osteonectin isolated from bovine tissue by Romberg et al. (1985) appears to be the same in size and composition as that initially described by Termine et al. (1981a). Romberg et al. (1985) report a molecular weight of 29 000, based upon sedimentation equilibrium analyses, compared to SDS–PAGE-generated apparent sizes of 32 000 and 40 000 on nonreduced and reduced gels, respectively. The effect of reduction on apparent size suggests that disulfide bonding significantly affects the conformation of the protein. Human osteonectin shows a similar shift in size based upon SDS–PAGE migration under nonreducing (M_r ~38 000) and reducing (M_r ~46 000) conditions (Fisher et al., 1987a). An

approximately 15-fold increase in iodoacetamide alkylation of reduced bovine and human osteonectin versus nonreduced is also consistent with the native molecule being highly disulfide bridged but is inconsistent with the cysteine content (3.5 Cys/286 amino acid residues) based upon amino acid analysis (Fisher et al., 1987).

Sage et al. (1984) have described an extracellular glycoprotein (43K) with properties very similar to bone osteonectin, first from cultured bovine aortic endothelial cells and also from cultured human foreskin and bovine ligament fibroblasts, several transformed murine cell lines, and bovine smooth muscle cells. Mason et al. (1986) have reported the deduced amino acid sequence (from cDNA sequences) for a secreted protein, acidic and rich in cysteine (SPARC) produced by developing mouse parietal endoderm and the transplantable Engelbreth–Holm–Swarm mouse tumor. Mouse SPARC is structurally and functionally similar to the 43K protein and osteonectin (Engel et al., 1987). The derived amino acid sequence for human placental SPARC (Swaroop et al., 1988) is identical with that of bone osteonectin reported in this paper.

Osteonectin is considered by some to be a bone-specific protein (Termine et al., 1981b; Jundt et al., 1987) and a possible indicator in serum for osteoblast function. Subsequent to initial studies using rabbit polyclonal antibodies to fetal bovine osteonectin, indicating the presence of osteonectin only in mineralized tissue (Termine et al., 1981b), several reports have appeared showing the presence and/or potential for biosynthesis of osteonectin in nonmineralized tissues and cells. Wasi et al. (1984) demonstrated cross-reactivity of the above antibodies with porcine periodontal ligament. They also isolated from this soft tissue a protein that appeared identical with osteonectin in regard to immunoreactivity, migration on SDS–PAGE gels, and binding to hydroxyapatite. Biosynthesis of proteins in confluent periodontal ligament fibroblasts resulted in immunoprecipitable protein that comigrated with bovine osteonectin on reduced and nonreduced SDS–PAGE gels (Wasi et al., 1984). Using bovine osteonectin cDNA for Northern analysis, Young et al. (1986) detected osteonectin mRNA in a variety of tissues. The strongest hybridization was seen with RNA from bone cells, an intermediate reaction for tendon, and a barely detectable level for brain and liver. Several cells in culture, including mouse, human, and chick embryo fibroblasts, rat osteosarcoma ROS17/2, and a chondrosarcoma cell line, all contained RNA hybridizable with the bovine osteonectin cDNA. The authors suggest, however, that the presence of the messenger RNA in cultured cells may be due to culture conditions or the cells being transformed. Stenner et al. (1986) have reported the presence of osteonectin in human platelets, using a solid-phase radioimmunoassay consisting in part of a mouse monoclonal antibovine osteonectin antibody.

The biological role of osteonectin is unknown. Most studies have focused upon its possible role in bone mineralization, initially because of its reported tissue-specific localization in bone and dentin (Termine et al., 1981a). Consistent with this hypothesis are in vitro biochemical studies. Bovine osteonectin is selectively adsorbed upon hydroxyapatite with a $K_d = 8 \times 10^{-8}$ M (Romberg et al., 1985). Osteonectin also specifically binds to degraded rat (Termine et al., 1981a) and bovine (Romberg et al., 1985) collagen. Intrinsic fluorescence studies have shown high-affinity binding ($K_d = 3 \times 10^{-7}$ M) of Ca^{2+} to osteonectin (Romberg et al., 1985). Termine and coworkers have also shown that osteonectin enhances the binding of hydroxyapatite (^{45}Ca labeled) to collagen as well as collagen-linked crystal formation from metastable solutions (Termine et al., 1981b), leading to the concept that osteonectin can physically mediate mineralization upon a collagen surface. Others (Romberg et al., 1985, 1986), however, have shown that osteonectin is a potent inhibitor in vitro of hydroxyapatite crystal growth. A 50% inhibition of crystal growth rate was obtained at 1.5×10^{-7} M osteonectin. This concentration is approximately one-fifth that required for similar inhibition by bone Gla protein (osteocalcin), thought to physiologically reduce bone mineralization (Price et al., 1982). It is also significantly lower on a molar basis than several known small-molecule inhibitors of crystal growth, including pyrophosphate, citrate, and Mg^{2+}. These data taken together suggest that osteonectin may play a direct role in bone and dental mineralization, but it is unclear as to the mechanism and as to whether osteonectin helps to increase or decrease mineralization. In vivo studies have not yielded a clear message regarding the relationship of osteonectin to bone mineralization. Termine's group in studies on *osteogenesis imperfecta* in cattle has reported decreased levels of bone osteonectin in one population (Termine et al., 1984) of affected individuals but not in another population (Fisher et al., 1986). Bone from humans with *osteogenesis imperfecta* appears to have a normal range of osteonectin concentration (Fisher et al., 1987b).

Reports of high levels of structural similarity with proteins from nonmineralized tissues of other species and the presence and synthesis of osteonectin in nonosteoid cells have raised the question as to whether osteonectin might be identical with SPARC and has additional or alternative functions in tissue modeling and repair and hemostasis. The results reported in this paper, taken together with those of Swaroop et al. (1988), clearly establish that the genetic material encoding osteonectin and SPARC is identical. Engel et al. (1987) have suggested several possible roles for SPARC in affecting cell activation and migration and the synthesis and modeling of the extracellular matrix during differentiation or tissue injury, that by analogy might also exist for osteonectin. Among these would be the inhibition of undesired extracellular mineralization in environments having high levels of Ca^{2+} and exposed collagen, either because of developmental or because of pathological stages. Another function, resulting from specific Ca^{2+} binding and consequential conformational changes, might be the regulation and/or direction of presently unknown biologically active proteins to the surface of cells, in a manner similar to that of the "Gla" domains in several of the protein blood clotting factors [see Suttie (1985, 1987)].

In the present study, two cDNA clones encoding for human osteonectin were isolated from independent SaOS-2 λgt11 libraries. The two cDNAs overlap one another by ~400 bp and have an identical sequence within this region. Furthermore, the SaOS-2 cDNA sequence is in complete agreement with the corresponding genomic material, indicating that the transcribed genetic material (and translated protein product) in the transformed SaOS-2 cell line is the same as that expected from normal cells.

Nucleotide analysis of the composite cDNA reveals the existence of a nascent protein consisting of a hydrophobic 17 amino acid signal peptide followed by a mature protein of 286 amino acids. From secondary structural characterization of human osteonectin, four general structural/functional domains may be assigned to the osteonectin molecule. Structurally, human osteonectin is identical with human placental SPARC and remarkably similar to their bovine (osteonectin) and mouse (SPARC) homologues. On the amino acid level, human osteonectin is 98% identical with bovine osteonectin with only

a single amino acid gap and 92% identical with mouse SPARC. Comparison of the signal peptide of human osteonectin with those predicted for bovine osteonectin and mouse SPARC reveals, for all 3 molecules, a leader sequence of 17 amino acids having 100% identity.

For domain I of the mature molecule, all three proteins are glutamic acid rich with no positively charged amino acids present. In addition, all three molecules have a duplicate internal repeat that resides within this section. However, this region of mouse SPARC does have a significantly greater degree of variability in sequence identity from human osteonectin (15 amino acid differences and 1 deletion; see Figure 5) and bovine osteonectin than in the remainder of the molecule.

Amino acid sequence and composition data (Fisher et al., 1987a) strongly suggest that there is no vitamin K dependent γ-carboxylation of glutamate in domain I of human osteonectin, even though an active carboxylation system exists in bone cells (Pan et al., 1984) and the protein sequence is very similar to those modified in the plasma clotting factors. It has been shown by several groups that the leader propeptide plays an important role in the γ-carboxylation of the clotting factors (Suttie et al., 1987; Jorgensen et al., 1987; Foster et al., 1987) and bone Gla protein (Pan & Price, 1985). Analysis of the osteonectin leader peptide shows no similarity to those of the clotting factors, and therefore may explain the lack of γ-carboxylation of osteonectin.

The second domain is cysteine, proline, and serine rich and has potential N-glycosylation sites. Two potential N-glycosylation sites (residues 71 and 99) exist in the human and bovine sequences whereas only one is found in mouse SPARC, further suggesting that the murine homologue may be more distantly related than human and bovine osteonectin. In all three species, the region shows homology to ovomucoid protease inhibitor domains. The functional significance of this region in osteonectin is unclear. Preliminary studies (R. Kelm and K. G. Mann, unpublished results) have failed to demonstrate coagulation protease inhibition by osteonectin. The high density of serine residues in this area makes it a likely candidate for serine phosphorylation. Potential serine phosphorylation sites (AspSerSer and AspSerGlu) identical with those reported for statherin (Schlesinger & Hay, 1977) are found in human osteonectin (residues 104–141), bovine osteonectin, and mouse SPARC. For human and bovine osteonectin, domain III contains a possible 12 amino acid "EF-hand" calcium binding region which is apparently not found in mouse SPARC. In common to all three molecules, the fourth domain contains the remaining cysteine residues and an "EF-hand" calcium binding structure. One notable difference between human and bovine osteonectin is the insertion of a cysteine residue at position 193 in the bovine molecule, thereby generating an additional potentially free sulfhydryl group.

A comparison of the nucleotide sequence reported in this paper and that reported by Swaroop et al. (1988) reveals only seven differences in the 3′-untranslated region. Nucleotide sequence identity of human osteonectin with bovine osteonectin and mouse SPARC is also extremely high, as shown in Figure 3, and shows that the protein's structure has been highly conserved. Percent nucleotide identity between human osteonectin and bovine osteonectin overall is 87% for the 5′ noncoding, 93% for the coding, and 76% for the 3′ noncoding. As expected, the most conserved region is that encoding the protein. At both the amino and nucleic acid sequence levels, the degree of species similarity between human and bovine osteonectin is much higher than for many other proteins. For comparison, 2 highly conserved proteins, plasma proteins C and S, have 74% (of 461 amino acid residues) and 82% (of 635 residues) identity, respectively, when comparing human to bovine (Beckman et al., 1985; Hoskins et al., 1987). Nucleotide sequence similarity between human osteonectin and mouse SPARC is also very high but not to the extent seen between human and bovine osteonectin. A comparison of nucleotide sequences for human osteonectin and mouse SPARC reveals 75% identity for the 5′-noncoding region, 90% for the coding region, and 61% for the 3′-noncoding region.

Using human–mouse somatic cell hybrids and Southern mapping, we have located the human osteonectin gene on chromosome 5 (Villarreal, Long, and Naylor, unpublished results). Southern hybridization of human genomic DNA with the 540 bp osteonectin cDNA from plasmid pHVON-9-2 yields a pattern (data not presented) that is consistent with the map shown in Figure 6, and is consistent with there being a single copy of the gene in the haploid genome. Swaroop et al. (1988) have independently recently reported the location of the human osteonectin gene on chromosome 5 and presented evidence for it being single-copy. Preliminary results of Findlay et al. (1988) for bovine osteonectin and of Hogan et al. (1988) for mouse SPARC indicate that single-copy genes exist in these organisms as well.

Gilbert (1978) originally proposed that intervening sequences (introns) in genes may separate distinct coding segments for protein structure/function domains, and thereby serve in the "shuffling" of common domains into related but different proteins. Examples consistent with this hypothesis are the vitamin K dependent plasma serine proteases involved in blood coagulation (Long, 1986), the globins (Maniatis et al., 1980), and ovomucoid (Stein et al., 1980). However, examples also exist where there is no apparent correlation between intron position and protein domains or conservation of intron position between homologous proteins, as exemplified by the ovalbumin/α_1-antitrypsin/antithrombin III family (Leicht et al., 1981; Prochownik et al., 1985). Traut (1988) has recently reviewed the size and structure/function relationships of eukaryotic exons. Consistent with his findings, all 9 of the coding exons for human osteonectin encode 20–50 amino acids, each of which can be considered as a potential minimal protein domain.

Our understanding of the exact functional components and structural elements for osteonectin, based largely on secondary structural predictions, is presently limited. All of the intron positions in the human osteonectin gene are located in positions identical with those for bovine osteonectin and mouse SPARC, indicating a remarkable conservation of genomic organization, and suggesting that they may denote important protein structure/function boundaries. With these points in mind, we have compared the proposed protein domains in human osteonectin with the gene's coding exons. Generally, there is good agreement with the proposed protein domains and the position of introns in the gene. The first coding exon contains all of the signal peptide and the first two amino acids of the mature protein, and thereby includes the signal peptide cleavage site. The second and third exons (amino acid residues 3–53) contain the strongly predicted α-helical, acidic acid-rich domain I of the mature protein. The fourth and fifth coding exons (residues 53–93 and 94–134) correlate well with the cysteine-rich domain II (residues 55–138). Each of the two exons contains a potential site for N-linked glycosylation. Introns D and F closely span the ovomucoid homologous region (residues 76–132) and are similar to those found bounding each of the chicken ovomucoid domains (Stein et al., 1980).

However, intron E, between amino acid residues 93 and 94, falls within the ovomucoid homologous domain at a position distant from that seen in the chicken ovomucoid gene (Stein et al., 1980). The sixth and seventh exons (residues 134–178 and 179–228) contain all of domain III (amino acids 138–212) and a small portion of domain IV. Domain IV (residues 213–286) is predominantly contained within the eighth and ninth exons (residues 228–278 and 278–286). The potential Ca^{2+} binding EF-hands (residues 165–176 and 257–268) are located at the carboxy terminii of exons 6 and 8, respectively.

At the present time in the case of osteonectin, a comparison of intron positions with proposed protein domains may be more important in further defining the protein domains than in testing the hypothesis that introns separate protein domains. A consideration of intron positions in the coding region of osteonectin can serve in directing and interpreting further physical and biochemical studies on the structure and function of osteonectin.

ACKNOWLEDGMENTS

We thank Drs. Mark Thiede, Mitchell Weiss, and Ross T. A. MacGillivray for providing human DNA libraries. The secretarial assistance of Jean Kenyon is acknowledged.

Registry No. DNA (human osteonectin mRNA complementary), 121210-40-0; osteonectin (human precursor reduced), 121210-48-8; osteonectin (human reduced), 121210-49-9.

REFERENCES

Beckman, R. J., Schmidt, R. J., Santerre, R. F., Plutzky, G., Crabtree, G. R., & Long, G. L. (1985) *Nucleic Acids Res. 13*, 5233–5247.

Birnboim, H. C., & Doly, J. (1979) *Nucleic Acids Res. 7*, 1513–1523.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., & Chambon, P. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 48–53.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45–148.

Davis, R. W., Botstein, D., & Roth, J. R. (1980) *Advanced Bacterial Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Engel, J., Taylor, W., Paulsson, M., Sage, H., & Hogan, B. (1987) *Biochemistry 26*, 6958–6965.

Feinberg, A. P., & Vogelstein, B. (1983) *Anal. Biochem. 132*, 6–13.

Findlay, D. M., Fisher, L. W., McQuillan, C. I., Termine, J. D., & Young, M. F. (1988) *Biochemistry 27*, 1483–1489.

Fisher, L. W., Denholm, L. J., Conn, K. M., & Termine, J. D. (1986) *Calcif. Tissue Int. 38*, 16–20.

Fisher, L. W., Hawkins, G. R., Tuross, N., & Termine, J. D. (1987a) *J. Biol. Chem. 262*, 9702–9708.

Fisher, L. W., Drum, M. A., Robey, P. G., Conn, K. M., & Termine, J. D. (1987b) *Calcif. Tissue Int. 40*, 260–264.

Foster, D. C., Rudinski, M. S., Schach, B. G., Berkner, K. L., Kumar, A. A., Hagen, F. S., Sprecher, C. A., Insley, M. Y., & Davie, E. W. (1987) *Biochemistry 26*, 7003–7011.

French, B. T., Maul, H. M., & Maul, G. G. (1986) *Anal. Biochem. 156*, 417–423.

Geddes, V. A., LeBonniec, B. F., Louie, G. V., Brayer, G. D., Thompson, A. R., & MacGillivray, R. T. A. (1989) *J. Biol. Chem. 264*, 4689–4697.

Gilbert, W. (1978) *Nature (London) 271*, 501.

Hoskins, J., Norman, D. K., Beckman, R. J., & Long, G. L. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 349–353.

Jorgensen, M. T., Cantor, A. B., Furie, B. C., Brown, C. L., Shoemaker, C. B., & Furie, B. (1987) *Cell 48*, 185–191.

Jundt, G., Berghauser, K.-H., Termine, J. D., & Schultz, A. (1987) *Cell Tissue Res. 248*, 409–415.

Katz, L., Kingsbury, P., & Hekinski, D. R. (1973) *J. Bacteriol. 114*, 577–591.

Kretsinger, R. H. (1980) *CRC Crit. Rev. Biochem. 8*, 119–174.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.

Laskowski, M., Ikunoshin, K., 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.

Leicht, M., Long, G. L., Chandra, T., Kurachi, K., Kidd, V. J., Mau, M., Jr., Davie, E. W., & Woo, S. L. C. (1982) *Nature (London) 297*, 655–660.

Long, G. L. (1986) *Cold Spring Harbor Symp. Quant. Biol. 51*, 525–529.

Maniatis, T., Fritsch, E. F., Lauer, J., & Lawn, R. M. (1980) *Annu. Rev. Genet. 14*, 145–178.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Mann, K., Deutzmann, R., Paulsson, M., & Timpl, R. (1987) *FEBS Lett. 218*, 167–172.

Mason, I. J., Taylor, A., Williams, J. G., Sage, H., & Hogan, B. L. M. (1986) *EMBO J. 5*, 1465–1472.

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.

McVey, J. H., Nomura, S., Kelly, P., Mason, I. J., & Hogan, B. L. M. (1988) *J. Biol. Chem. 263*, 11111–11116.

Mount, S. W. (1982) *Nucleic Acids Res. 10*, 459–472.

Ogden, R. C., & Adams, D. A. (1987) *Methods Enzymol. 152*, 61–87.

Pan, L. C., & Price, P. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6109–6113.

Pan, L. C., Williamson, M. K., & Price, P. A. (1985) *J. Biol. Chem. 260*, 13398–13401.

Price, P. A., Williamson, M. K., Haba, T., Dell, R. B., & Jee, W. S. S. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 7734–7738.

Prochownik, E. V., Bock, S. C., & Orkin, S. H. (1985) *J. Biol. Chem. 260*, 9608–9612.

Romberg, R. W., Werness, P. G., Lollar, P., Riggs, B. L., & Mann, K. G. (1985) *J. Biol. Chem. 260*, 2728–2736.

Romberg, R. W., Werness, P. G., Riggs, B. L., & Mann, K. G. (1986) *Biochemistry 25*, 1176–1180.

Sage, H., Johnson, C., & Bornstein, P. (1984) *J. Biol. Chem. 259*, 3993–4007.

Sage, H., Tupper, J., & Bramson, R. (1986) *J. Cell. Physiol. 127*, 373–387.

Schlesinger, D. H., & Hay, D. I. (1977) *J. Biol. Chem. 252*, 1689–1695.

Sharp, P. A. (1987) *Science 235*, 766–771.

Stein, J. P., Catterall, J. F., Kristo, P., Means, A. R., & O'Malley, B. W. (1980) *Cell 21*, 681–687.

Stenner, D. D., Romberg, R. W., Tracy, R. P., Katzmann, J. A., Riggs, B. L., & Mann, K. G. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2868–2872.

Stenner, D. D., Tracy, R. P., Riggs, B. L., & Mann, K. G. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6892–6896.

Suttie, J. W. (1985) *Annu. Rev. Biochem. 54*, 459–477.
Suttie, J. W. (1987) *Current Advances in Vitamin K Research*, Elsevier, New York.
Suttie, J. W., Hoskins, J. A., Engelke, J., Hopfgartner, A., Ehrlich, H., Bang, N. U., Belagaje, R. M., Schoner, B., & Long, G. L. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 634–637.
Swaroop, A., Hogan, B. L. M., & Francke, U. (1988) *Genomics 2*, 37–47.
Termine, J. D., Kleinman, H. K., Whitson, S. W., Conn, K. M., McGarvey, M. L., & Martin, G. R. (1981a) *Cell 26*, 99–105.
Termine, J. D., Belcourt, A. B., Conn, K. M., & Kleinman, H. K. (1981b) *J. Biol. Chem. 256*, 10403–10408.
Termine, J. D., Robey, P. G., Fisher, L. W., Shimokawa, H., Drum, M. A., Conn, K. M., Hawkins, G. R., Cruz, J. B., & Thompson, K. G. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2213–2217.
Tracy, R. P., Shull, S., Riggs, B. L., & Mann, K. G. (1988) *Int. J. Biochem. 20*, 653–660.
Traut, T. W. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2944–2948.
Wasi, S., Otuska, K., Yao, K.-L., Tung, P. S., Aubin, J. E., Sodek, J., & Termine, J. D. (1984) *Can. J. Biochem. Cell Biol. 62*, 470–478.
Young, R. A., & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1194–1198.
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.
Zhang, H., Scholl, R., Browse, J., & Somerville, C. (1988) *Nucleic Acids Res. 16*, 1220.

Separation of the Complex Asparagine-Linked Oligosaccharides of the Glycoprotein Fetuin and Elucidation of Three Triantennary Structures Having Sialic Acids Linked Only to Galactose Residues†

Brad Bendiak,*,‡ Marees Harris-Brandts,§ Stephen W. Michnick,§ Jeremy P. Carver,§ and Dale A. Cumming§,‖
Department of Biochemistry, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8, and Departments of Medical Genetics and Medical Biophysics, University of Toronto, Toronto, Ontario, Canada M5S 1A8
Received October 12, 1988; Revised Manuscript Received April 18, 1989

ABSTRACT: Asparagine-linked oligosaccharides of the glycoprotein fetuin were isolated as reducing oligosaccharides after hydrazinolysis/re-N-acetylation/mild acid treatment of the Pronase-digested glycoprotein. The sialylated oligosaccharides were separated by high-performance liquid chromatography in two different systems, which resulted in greater than 35 fractions, comprising di-, tri-, tetra-, and pentasialylated oligosaccharides. The major components were isomeric structures comprising the tri- and tetrasialylated fractions. In this and the accompanying paper (Cumming et al., 1989), the structures of 10 of the major components of the tri-, tetra-, and pentasialylated oligosaccharide fractions are described. Separation protocols and three isolated structures having sialic acid linked only to galactose are presented in this paper.

Oligosaccharides of the glycoprotein fetuin have been investigated for over 25 years (Spiro, 1962, 1964). As such, attempts to elucidate their structures have spanned virtually the entire time period of glycoprotein structural research and have generally reflected the history of technological advances achieved during development of the field. Initial investigations were performed with glycopeptide mixtures of proteolytically treated fetuin. Spiro and colleagues have provided good evidence for the structures of the major (Spiro & Bhoyroo, 1974) and minor (Edge & Spiro, 1987) O-linked oligosaccharides of fetuin, after release of the oligosaccharides from the peptide by an alkaline β-elimination. They also elucidated the general features of the N-linked structures, i.e., a branched oligosaccharide having Man and GlcNAc residues in the core region, with peripheral GlcNAc, Gal, and NeuAc residues extending from the core (Spiro, 1964).

Two conflicting reports of the Asn-linked structures appeared later (Nilsson et al., 1979; Baenziger & Fiete, 1979). Again, glycopeptides were prepared from the intact protein; no reports were made of successful subfractionation of the Asn-linked oligosaccharides in these papers, possibly due to the difficulties in separating mixtures of glycopeptides that vary in both the oligosaccharide and peptide components. At this time, neither high-field NMR[1] (Dorland et al., 1977; Strecker et al., 1977; Carver & Grey, 1981) nor HPLC was in common use for oligosaccharide structural determination.

More recently, it has been demonstrated that heterogeneity exists in the Asn-linked structures of fetuin, particularly in the presence of a Galβ1→3GlcNAc in some of the oligosaccharides (Berman, 1986; Townsend et al., 1986; Berman et al., 1988). These studies were performed with mixtures of glycopeptides, and results were deduced from the data of a mixture of oligosaccharide structures. Of importance, Takasaki and Kobata (1986) employed hydrazinolysis to liberate

† We thank the MRC of Canada and the Terry Fox Special Initiatives Fund for support. B.B. thanks Harry Schachter for MRC support and Glenn Carnegie for acquisition of a number of the NMR spectra.
* Address correspondence to this author at The Biomembrane Institute, Suite 305, 201 Elliott Ave. West, Seattle, WA 98119.
‡ The Hospital for Sick Children.
§ University of Toronto.
‖ Present address: Genetics Institute, 87 Cambridge Park Dr., Cambridge, MA 02140.

[1] Abbreviations: NMR, nuclear magnetic resonance spectroscopy; HPLC, high-performance liquid chromatography; DEAE, 2-(diethylamino)ethyl; Tris, tris(hydroxymethyl)aminomethane; NOE, nuclear Overhauser enhancement.

0006-2960/89/0428-6491$01.50/0 © 1989 American Chemical Society