Serum Vitamin D-binding Protein is a Third Member of the Albumin and Alpha Fetoprotein Gene Family

Nancy E. Cooke
Departments of Medicine and Human Genetics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104

E. Vivek David
Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104

Abstract

A near full-length cDNA encoding the human vitamin D-binding protein (hDBP) was isolated from a human liver mRNA expression library. Complete sequence analysis of this clone predicts the full-length amino acid sequence of the pre-hDBP. Comparison of the sequence of the hDBP mRNA and protein to existing protein and nucleic acid data banks demonstrates a strong and highly characteristic homology of the hDBP with human albumin (hALB) and human α-fetoprotein (hAFP). Based upon this structural comparison, we establish that DBP is a member of the ALB and AFP gene family.

Introduction

Human vitamin D-binding protein (hDBP),[1] also known as Gc-globulin, is an abundant, multifunctional, and highly polymorphic serum glycoprotein synthesized by the liver (1). DBP is the major serum transport protein for the vitamin D sterols (2), binds and sequesters monomers of actin with high affinity (3, 4), and has been identified on the surface of a variety of cell types including B-lymphocytes (5), subpopulations of T-lymphocytes (6), and the cytotrophoblasts of the placenta (7). The physiologic importance of these functions and their possible interrelationships remains to be defined. As an initial step in the detailed study of this abundant serum protein, we have isolated a near full-length copy (cDNA) of the hDBP mRNA, determined the primary structure of the encoded protein, and detected a close evolutionary and genetic relationship between DBP and two other abundant serum proteins, albumin (ALB) and α-fetoprotein (AFP).

Methods

RNA isolation and gel analysis. Total cellular RNA was isolated by standard technique (8) from human liver obtained at autopsy, from fresh rat liver (Sprague-Dawley), and from Hep 3B cells (9) grown in monolayer culture. PolyA+ mRNA was isolated from the total RNA by oligo-dT chromtography (10). The mRNAs were size-fractionated by electrophoresis through a 1% agarose-formaldehyde gel (11), transferred to nitrocellulose paper (12), and hybridized at reduced stringency (37°C with other conditions, as described in reference 12) to a nick-translated ^{32}P-labeled cDNA probe (13). After overnight hybridization, the filter was washed (at 42°C, otherwise as described in reference 12) and autoradiographed at −70°C on XAR film (Eastman Kodak Co., Rochester, NY) in the presence of a Cronex Lightening Plus intensifying screen. (E. I. DuPont de Nemours, Wilmington, DE). Ribosomal RNA from human reticulocytes (28S and 18S) and *Escherichia coli* (23S and 16S) were run on the gel in separate lanes, visualized by ethidium bromide staining, and used as molecular weight markers.

cDNA library screening. A library of human liver cDNA inserted in the expression vector λgt11 (14) was a generous gift of S. Woo, Howard Hughes Medical Institute, Baylor College of Medicine, Texas Medical Center, Houston, TX (15). This library was plated out on Y1090 bacteria and the plaques were screened by the method of Benton and Davis (16). All positives were plaque-purified by serial platings at low dilution. cDNA inserts were restriction mapped from phage DNA isolated by a mini-lysates procedure (17).

cDNA subcloning. DNA from each phage mini-lysate was digested with the restriction endonuclease EcoRI (New England Biolabs, Beverly, MA) at 1 U/μg DNA under conditions suggested by the manufacturer.

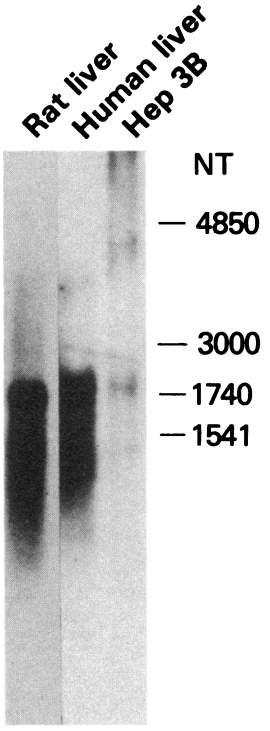

Figure 1. Detection of the human DBP mRNA by cross hybridization with a rat DBP cDNA probe. 5 μg of polyA+ mRNA from rat liver, 5 μg of polyA+ mRNA from human liver, or 20 μg of total RNA from Hep 3B cells were electrophoresed through a 1% agarose-formaldehyde gel, transferred to nitrocellulose paper, and hybridized at reduced stringency to a nick-translated, ^{32}P-labeled rDBP cDNA probe (Cooke, N. E., unpublished data). An autoradiograph of the hybridized filter is shown with the lanes identified. The position of ribosomal RNA molecular size markers are indicated in nucleotides (NT).

Address reprint requests to Dr. Nancy E. Cooke, Room 531, Johnson Pavilion/G2, University of Pennsylvania School of Medicine, Philadelphia, PA 19104.

Received for publication 23 September 1985.

1. *Abbreviations used in this paper:* AFP, α-fetoprotein; ALB, albumin; hDBP, human vitamin D-binding protein.

J. Clin. Invest.

0021-9738/85/12/2420/05 $1.00
Volume 76, December 1985, 2420–2424

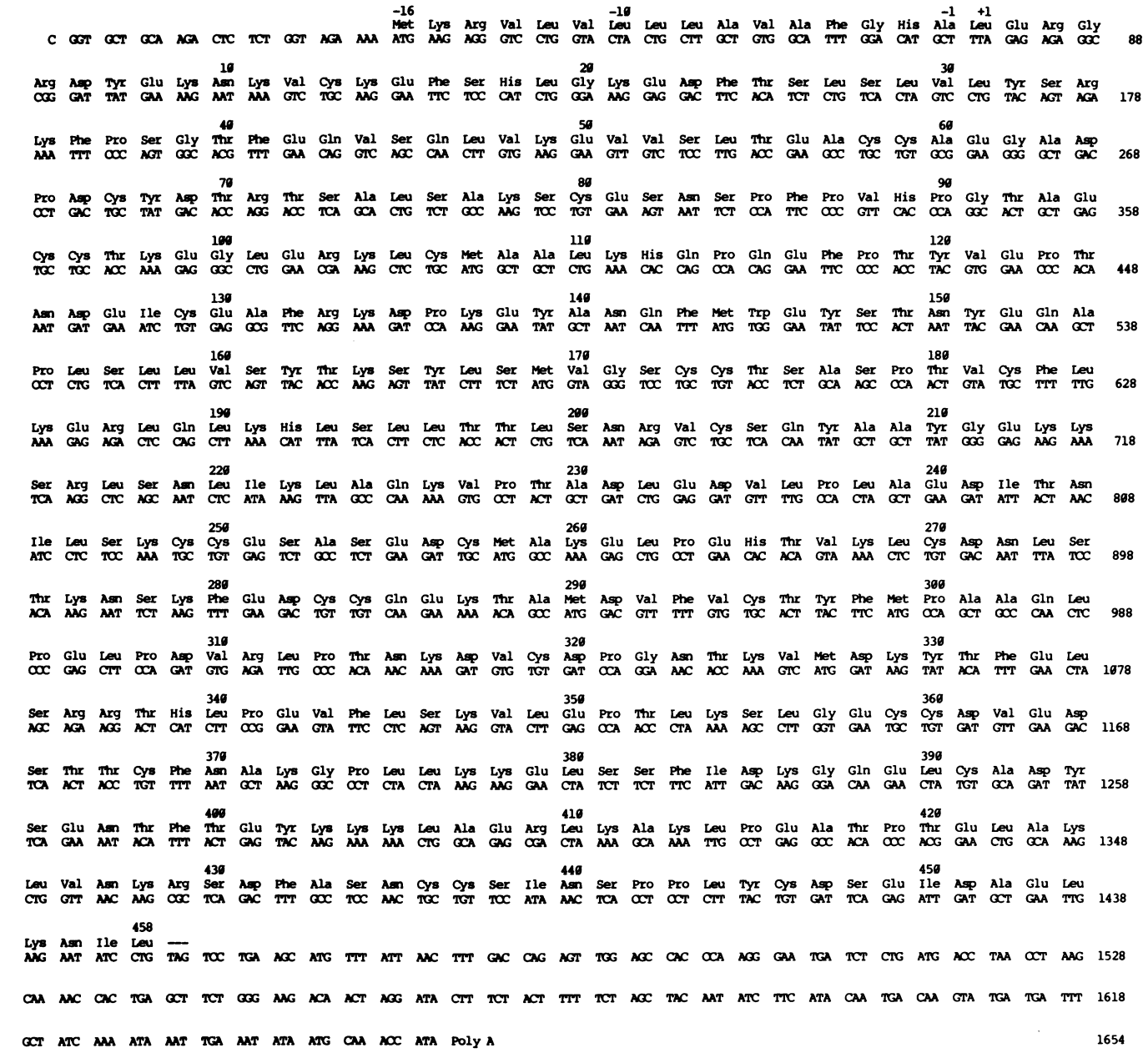

Figure 2. The nucleotide sequence of the hDBP cDNA and the full-length predicted amino acid sequence of hDBP. Nucleotides are numbered beginning with the first base of the hDBP cDNA insert. Amino acids are numbered with 1 representing the first residue of the mature protein.

The digest was phenol-extracted and ligated to EcoRI-digested and dephosphorylated SP65 plasmid (18) at a 5:1 molar ratio of insert to vector using 400 U of T4 ligase (New England Biolabs) in a total volume of 20 μl containing 50 mM Tris HCl, pH 7.4, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM spermidine, and 1 mM ATP. The entire ligation reaction was used to transform *E. coli* HB101 rendered competent by cold $CaCl_2$ treatment (19). Transformations were spread on L-broth plates containing ampicillin (35 μg/μl). Ampicillin-resistant transformants were screened in situ (20) with ^{32}P-labeled cDNA.

Plasmid preparation. All recombinant work was done under P1 containment conditions. Bacterial plasmids were grown in 1-liter batches to an OD_{660} of 0.8, amplified with chloramphenicol (170 μg/μl), and harvested 18 h later. Supercoiled plasmid was purified from a clarified lysate by cesium chloride isopycnic centrifugation in the presence of 0.75 mg/ml ethidium bromide, harvested, and phenol-extracted before use.

DNA sequencing. Restriction sites with 5′ protruding ends were dephosphorylated with calf intestinal alkaline phosphatase (Bethesda Research Laboratories, Gaithersburg, MD) and ^{32}P end-labeled in a 25-μl reaction with T4 polynucleotide kinase (New England Biolabs) and (^{32}P)rATP (5,000 Ci/mM, Amersham Corp., Arlington Heights, IL). 3′ recessed ends were labeled with *E. coli* DNA polymerase, Klenow fragment (Bethesda Research Laboratories), and the appropriate (^{32}P)dNTP in the recommended buffer. End-labeled fragments were either strand-separated (21) or digested with a second restriction enzyme before gel purification and sequencing. All DNA sequencing was by the method of Maxam and Gilbert (22) with modifications as previously described (23). All regions were either sequenced on both strands, or two separate times on the same strand.

Primer extension analysis. For primer extension analysis a gel-purified end-labeled ^{32}P cDNA fragment was hybridized to liver mRNA in 30 μl

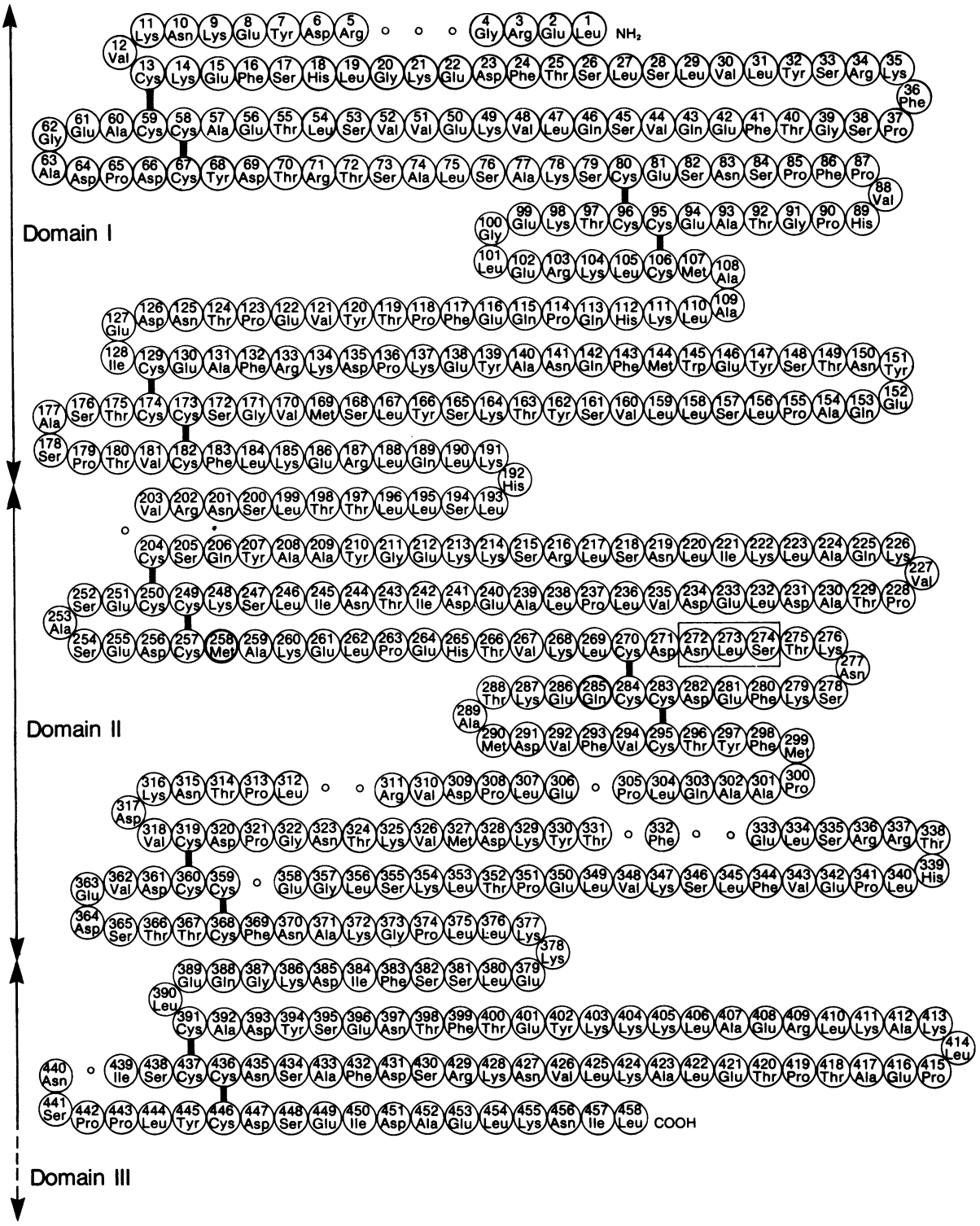

Figure 3. Conservation of cysteine positioning in hDBP and hALB. The predicted amino acid sequence of hDBP is displayed in the format originally proposed by J. R. Brown for albumin (30). Deletions in hDBP, when aligned with hALB, are indicated by the small circles, and insertions by bold outline. A potential N-linked glycosylation site is enclosed in a rectangle. The position of the homologous internal domains is indicated in the left margin.

of hybridization buffer (80% deionized formamide, 0.4 M NaCl, 40 mM PIPES, pH 6.4, 1 mM EDTA) at 42°C for 3 h. Primer extension with avian myeloblastosis virus reverse transcriptase was subsequently carried out as previously described (24) with no significant modification.

Computer analysis. GenBank, European Molecular Biology Labs, and National Biomedical Research Foundation databases and software used in the data analysis were accessed through the Bionet National Computer Resource for Molecular Biology, Palo Alto, CA.

Results

A rat DBP cDNA clone (Cooke, N. E., unpublished data), selected by an immunological approach (to be described elsewhere), was shown by Northern analysis to cross-hybridize under conditions of reduced stringency to a single mRNA species of 1,750 nucleotides from the human liver (Fig. 1). This same mRNA is present in the human hepatoma cell line Hep 3B which is known to synthesize DBP (25) but not in human placental mRNA (data not shown). A slightly smaller mRNA species of 1,650 nucleotides is seen in the rat liver mRNA. Utilizing this low stringency cross-hybridization between the rDBP cDNA and the hDBP mRNA, several human liver DBP cDNA clones were selected from a human liver cDNA library constructed in the bacteriophage λgt11 (15). The cDNA insert in each of the hybridization-positive phage clones was mapped, and the largest, 1,383 nucleotides long, was fully restriction-mapped and sequenced. The identity of this clone as encoding hDBP was established by identification of three previously sequenced peptide fragments of hDBP (amino acids 1–41 [26], 414–419, and 421–424 [27]) and by 77% homology to the terminal 441 amino acids of the rDBP cDNA clone. Using the 5′ EcoRI fragment of this partial clone as probe, an overlapping hDBP cDNA clone extending further in the 5′ direction was selected and restriction-mapped. This recombinant contained a 1,654 base-pair cDNA insert. The sequence of this human cDNA, shown in Fig. 2, contains a 1,422 nucleotide open reading frame beginning with a Met codon at the 29th base.

To establish whether the cDNA displayed in Fig. 2 was full-length, a 107 base-pair restriction fragment labeled at the EcoRI site (nucleotide 119) and extending to the HinfI site (nucleotide 12) was isolated from the 5′ end of the cDNA, and hybridized to human liver RNA. cDNA synthesis was extended with reverse transcriptase. The resultant cDNA, sized next to a DNA sequence ladder, extended 48 nucleotides 5′ to the labeled fragment (data not shown). Thus the reported cDNA sequence begins 36 base pairs 3′ to the mRNA cap site and the full-length hDBP mRNA is predicted to contain 1,690 nucleotides, excluding its polyA$^+$ tail.

Discussion

The full-length primary sequence of hDBP, as displayed in Fig. 2, can be correlated with several of the known physical properties of this protein. DBP is a secreted protein and would therefore be predicted to have a signal sequence. The amino-terminus of the mature serum protein was established by aligning the sequence predicted by the cDNA to a sequenced amino-terminal hDBP peptide (27). From this alignment we deduce that the coding region of prehDBP begins with a 16-amino-acid hydrophobic signal sequence. Human DBP is known to be glycosylated and the determined sequence contains one potential N-linked glycosylation site (Fig. 3, enclosed in rectangle). The encoded mature hDBP contains 458 residues with a calculated molecular weight of 51,335, compared with experimental estimates of 54,000–58,000 for the serum glycoprotein.

DBP, ALB, and the fetal analogue of ALB, AFP, are major serum proteins synthesized by mammalian liver parenchymal cells. All three molecules possess a high content of cysteine residues. Comparison of the cysteines in hDBP to those in hALB and hAFP reveals a high degree of conservation in their number and positions. An alignment of hDBP, hALB (28), and hAFP (29) using the cysteine residues as landmarks, revealed significant additional amino acid and nucleotide homologies (Table I). The similarities between these three molecules are so strong that hDBP can be easily displayed in the format proposed by J. R. Brown (30) for albumin (Fig. 3). Human DBP terminates 123 amino acids before hALB or hAFP, and there is a rapid drift in the 3′ untranslated region of hDBP compared with the continued coding regions of the other two mRNAs. Human DBP contains a unique tryptophan at position 145 not present in hALB, or hAFP. The glycosylation pattern of the three proteins appears to differ: the predicted glycosylation sites in hAFP are not conserved in hDBP, and hALB is not glycosylated at all.

The triplicated internal domain structure first noted in ALB and AFP is also present in hDBP (Table I) with the exception of the previously mentioned truncation within the third domain of hDBP. This suggests a common evolutionary origin of these three serum proteins from a precursor with a triplication of an original single domain. The hDBP gene may in fact be situated in close proximity to the hALB and hAFP genes. Previous studies of families with albumin mutations have documented a linkage of <1.5 centimorgan map units between the hDBP and the hALB loci on human chromosome 4 in segment 4q11–13 (31–34). This localization and linkage has been recently confirmed by in situ hybridization using the hDBP cDNA as probe (Cooke, N. E., H. Willard, E. V. David, and D. George, unpublished

Table I. Amino Acid and Nucleotide Sequence Comparisons between hDBP, hALB, hAFP, and Internal Domains

	Amino acids						Coding nucleotides		
		Identical		Conservative replacement (35)				Homology	
Comparison	Gaps	No.	Percent	No.	Percent	Total* percent	Gaps	No.	Percent
DBP/ALB (28)	14	112/487	23.0	124/487	25.4	48.5	42	570/1,461	39.0
DBP/AFP (29)	16	95/448	19.5	118/488	24.2	43.7	48	536/1,464	36.6
ALB/AFP	12	244/616	39.6	160/616	26.0	65.6	36	953/1,848	51.6
DBP internal domains									
I/II	10	45/194	23.2	43/194	22.2	45.4	30	224/582	38.5
I/III	1	12/81	14.8	20/81	24.6	39.4	3	79/243	32.5
II/III	3	15/81	18.5	22/81	27.2	45.7	9	89/243	36.6

Gaps were inserted to maximize amino acid homology. The signal peptides are included in the amino acid analyses. Total lengths include gaps which are treated as mismatches. * Total includes identical and conservative replacements.

data). Based upon the primary structural homologies at both mRNA and protein levels, the presumed similarities in the secondary structures indicated by strictly conserved placement of cysteine residues, and the linkage of the three encoding genes on chromosome 4, we conclude that DBP is a member of the gene family that encodes the other major serum proteins, ALB and AFP.

Note added in proof. While this manuscript was in press, we learned of the work of Yang et al. (36) in which a DBP cDNA of the Gc^2 type was characterized. Our sequence differs from theirs in positions 152, 311, 416, and 420. The latter two amino acids of our sequence are consistent with those reported in Gc by Svasti et al. (27) and suggest that our sequence represents the Gc^1 allele.

Acknowledgments

This work was supported by National Institutes of Health research grant RO1GM32035 (Dr. Cooke). Dr. Cooke is a recipient of a Basil O'Connor Starter Grant from the National Foundation–March of Dimes.

References

1. Haddad, J. G. 1979. Transport of vitamin D metabolites. *Clin. Orthop. Relat. Res.* 142:249–274.

2. Daiger, S. P., M. S. Schanfield, and L. L. Cavalli-Sforza. 1975. Group specific component (Gc) proteins bind vitamin D and 25-hydroxyvitamin D. *Proc. Natl. Acad. Sci. USA.* 72:2076–2080.

3. Cooke, N. E., J. Walgate, and J. G. Haddad. 1979. Human serum binding protein for vitamin D and its metabolites. I. Physicochemical and immunological identification in human tissues. *J. Biol. Chem.* 254: 5965–5971.

4. Van Baelen, H., R. Bouillon, and P. DeMoor. 1980. Vitamin D binding protein (Gc-Globulin) binds actin. *J. Biol. Chem.* 255:2270–2272.

5. Petrini, M., D. L. Emerson, and R. M. Galbraith. 1983. Linkage between surface immunoglobulin and cytoskeleton of B lymphocytes may involve Gc protein. *Nature (Lond.).* 306:73–74.

6. Petrini, M., R. M. Galbraith, D. L. Emerson, A. E. Nel, and P. Arnaud. 1985. Structural studies of T lymphocyte Fc receptors. *J. Biol. Chem.* 260:1804–1810.

7. Nestler, J. E., E. Lange, J. McLeod, H. J. Kliman, and J. F. Strauss. 1985. Detection of vitamin D binding protein (DBP) in human cytotrophoblasts. *Seventh Annual American Society for Bone and Mineral Research.* Wash. DC. 266*a*. (Abstr.).

8. Chirgwin, J., A. E. Przybyla, R. S. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry.* 18:5294–5301.

9. Knowles, B. B., C. Howe, and D. P. Aden. 1980. Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. *Science (Wash. DC).* 209:497–499.

10. Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA.* 69:1408–1412.

11. Rave, N., R. Crkvenjakov, and H. Boedtker. 1979. Identification of procollagen mRNAs transferred to diazobenzyloxymethyl paper from formaldehyde agarose gels. *Nucl. Acids Res.* 6:3559–3567.

12. Thomas, P. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA.* 77:5201–5205.

13. Maniatis, T., A. Jeffrey, and D. G. Kleid. 1975. Nucleotide sequence of the rightward operator of Phage λ. *Proc. Natl. Acad. Sci. USA.* 72:1184–1188.

14. Young, R. A., and R. W. Davis. 1983. Yeast RNA polymerase II genes: isolation with antibody probes. *Science (Wash. DC).* 222:778–782.

15. Kwok, S. C. M., F. D. Ledley, A. G. DiLella, K. J. H. Robson, and S. L. C. Woo. 1985. Nucleotide sequence of a full-length complementary DNA clone and amino acid sequence of human phenylalanine hydroxylase. *Biochemistry.* 24:556–561.

16. Benton, W. D., and R. W. Davis. 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. *Science (Wash. DC).* 196:180–182.

17. Leder, P., D. Tiemeier, and L. Enquist. 1977. EK2 derivatives of bacteriophage lambda useful in the cloning of DNA from higher organisms: the λgtWES system. *Science (Wash. DC).* 196:175–177.

18. Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucl. Acids Res.* 12:7035–7054.

19. Dagert, M., and S. D. Ehrlich. 1979. Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. *Gene (Amst.).* 6:23–28.

20. Grunstein, M., and D. Hogness. 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA.* 72:3961–3965.

21. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific cleavage reactions. *Methods Enzymol.* 65:499–560.

22. Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA.* 74:560–564.

23. Cooke, N. E., D. Coit, R. I. Weiner, J. D. Baxter, and J. A. Martial. 1980. Structure of the cloned DNA complementary to rat prolactin messenger RNA. *J. Biol. Chem.* 255:6502–6510.

24. Liebhaber, S. A., and Y. W. Kan. 1981. Differentiation of the mRNA transcripts originating from the α1 and α2 globin loci in normals and α-thalassemics. *J. Clin. Invest.* 68:439–446.

25. Haddad, J. G., D. P. Aden, and M. A. Kowalski. 1983. Characterization of the human plasma binding protein for vitamin D and its metabolites synthesized by the hepatoma-derived cell line, Hep 3B. *J. Biol. Chem.* 258:6850–6854.

26. Schoentgen, F., M. H. Metz-Boutigue, J. Jolles, J. Constans, and P. Jolles. 1985. Homology between the human vitamin D-binding protein (group specific component), α-fetoprotein and serum albumin. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 185:47–50.

27. Svasti, J., A. Kurosky, A. Bennett, and B. H. Bowman. 1979. Molecular basis for the three major forms of human serum vitamin D binding protein (group specific component). *Biochemistry.* 18:1611–1617.

28. Dugaiczyk, A., S. W. Law, and O. E. Dennison. 1982. Nucleotide sequence and the encoded amino acids of human serum albumin mRNA. *Proc. Natl. Acad. Sci. USA.* 79:71–75.

29. Morinaga, T., S. Masaharu, T. G. Wegman, and T. Tamaoki. 1983. Primary structures of human α-fetoprotein and its mRNA. *Proc. Natl. Acad. Sci. USA.* 80:4604–4608.

30. Brown, J. R. 1976. Structural origins of mammalian albumin. *Fed. Proc.* 35:2141–2144.

31. Weitkamp, L. R., D. L. Rucknagel, and H. Gershowitz. 1966. Genetic linkage between structural loci for albumin and group specific component (Gc). *Am. J. Hum. Genet.* 18:559–566.

32. Kaarsalo, E., L. Melartin, and B. S. Blumberg. 1967. Autosomal linkage between the albumin and Gc loci in humans. *Science (Wash. DC).* 158:123–125.

33. Mikkelson, M., P. Jacobsen, and K. Henningsen. 1977. Possible localization of Gc-system on chromosome 4. *Hum. Hered.* 27:105–107.

34. Harper, M. E., and A. Dugaiczyk. 1983. Linkage of the evolutionarily-related serum albumin and α-fetoprotein genes within q11–22 of human chromosome 4. *Am. J. Hum. Genet.* 35:565–572.

35. Dayhoff, M. O. 1978. Atlas of Protein Sequence. Vol. 5, Suppl. 3. National Biomedical Research Foundation, Washington, DC. 345–352.

36. Yang, F., J. L. Brune, S. L. Naylor, R. L. Cupples, K. H. Naberhaus, and B. H. Bowman. 1985. Human group-specific component (Gc) demonstrates striking homology with the albumin family of genes. *Proc. Natl. Acad. Sci. USA.* In press.