The complete rat vitamin D binding protein (DBP) gene has been cloned and characterized. Genomic mapping suggests that there is only one copy of this gene in the haploid genome. The gene spans 35 kilobase pairs and contains 13 exons. All exons, exon/intron borders, and 2196 base pairs of 5'-flanking region have been sequenced. The transcription start site, determined by primer extension analysis, is 62 base pairs upstream from the start of translation and predicts that an unusual TGTAAA motif may serve as a surrogate TATA. The promoter region contains about 50% nucleotide sequence similarity to the corresponding region of the partially characterized human DBP gene and is uniquely interrupted by a repetitive element. Although lacking in overall sequence similarity to the albumin (ALB) and a-fetoprotein (AFP) genes, the 5'-flanking region of the DBP gene contains a number of conserved segments which may correspond to critical proximal promoter elements in this gene family. The location of the introns in the coding region of the DBP gene is highly conserved when compared with the ALB and AFP genes. Detailed comparison of exon size and content confirms the previous prediction that the smaller size of the DBP protein results from loss of internal exons 12 and 13 from the DBP progenitor gene during its evolutionary divergence from ALB and AFP.

Vitamin D binding protein (DBP) is a major plasma constituent identified in all mammalian and avian species thus far examined (1). DBP is a monomeric, multifunctional glycoprotein with a molecular size of 58 kDa in humans (2) and 52 kDa in rat (3). It was originally identified in 1965 as a polymorphic serum protein, designated as the group-specific component of serum or Gc-globulin (4). It is the major plasma carrier protein for vitamin D and vitamin D metabolites containing one sterol binding site/molecule (dissociation constant for binding to 25(OH)D3, $K_d = 5 \times 10^{-9}$) (5). DBP is present in marked excess (5–8 µM) over that necessary for its role in vitamin D transport (0.1–0.2 µM), suggesting that it may have additional roles (6). Also, DBP like gelsolin has been found to be a serum actin binding protein. Monomeric actin binds to DBP with an affinity ($K_d = 1$ nM) higher than its affinity for the elongating actin filament (7). Therefore, DBP is capable of sequestering monomeric actin away from polymerization (8, 9). This has been proposed to constitute an effective mechanism whereby actin, released to the circulation during cell destruction, could be prevented from polymerization (10, 11). Neutrophil chemotaxis was found to be increased when DBP binds to C5a and C5a des-Arg, suggesting a "cochemotaxin" role for DBP (12). DBP also binds unsaturated fatty acids, although not as avidly as ALB. DBP is associated with the immunoglobulin Fc receptor (13) and the surface membranes of immunocytes (14, 15), cytotoxic blasts (16), and the transformed human monocyte-derived cell line U937. However, the inability to detect DBP mRNA in any of these cells suggests that cell-associated DBP may be exogenously derived from plasma (17). An extensive worldwide screening of human populations has failed to identify a Gco homozygote in spite of detection of a Gco allele (18). These findings have led to the suggestion that complete deficiency of DBP may be a lethal mutation and that the DBP gene might encode a function, known or as yet undiscovered, necessary to life (19).

Based on mRNA and protein sequencing data, DBP is postulated to be a member of a multigene family that includes ALB and AFP (20–22). Several lines of evidence suggest that this family arose from a triplication of a primordial ancestral gene. From sequence comparisons, it appears that an initial duplication resulted in DBP and the immediate ALB/AFP precursor gene, and a later reduplication, about 300–500 million years ago, resulted in the ALB and AFP genes (23, 24). The association of the three genes on human chromosome 4 (22, 25) and on rat chromosome 14 (26) supports their common origin. All three proteins share a series of highly conserved cysteine residues conferring a similar secondary folding structure with three internally homologous domains. However, unlike ALB and AFP, the third domain of DBP is truncated by 124 amino acids (21). While the structural organization of the ALB gene has been elucidated for human (27) and rat (28, 29) and that of AFP for human (30), rat (31–33), and mouse (23, 34), the structure of the DBP gene is not known. However, based on nucleotide and predicted amino acid sequence similarity between the cDNAs encoding ALB, AFP, and DBP, putative intron/exon junctions within the DBP gene were assigned (21, 35). From such sequence alignments, it was predicted that the truncation of DBP relative to ALB and AFP resulted from deletion of two internal exons in the third domain of a DBP progenitor gene rather than a premature termination mutation (35).
The ALB and AFP genes have served as major models for tissue-specific and developmentally regulated gene expression (36-40). Consequently, much detailed data has accumulated on the structure of their proximal and distal promoters (33, 41-58), while little information is available for DBP. Surprisingly, in a limited report of the human (h) DBP gene, which included 1649 bp of 5'-flanking sequence, exon 1, and 37 bp of intron 1, no similarity between its promoter region and that of ALB or AFP was found (59). We report here the overall structure of the rDBP gene as well as a complete sequence analysis of its exon/intron boundaries and 2.19 kb of its 5'-flanking region. Based on this information, we examine predictions made regarding its structure and compare its 5'-flanking region with those of AFP and ALB. These studies provide the basis for a more complete understanding of the molecular evolution of the DBP gene and its relationship to the ALB/ AFP/DBP multigene family.

**EXPERIMENTAL PROCEDURES**

**Materials and General Methods—**Except where noted otherwise, standard techniques were used for the isolation and manipulation of DNA and RNA (60). Restriction and modification enzymes were purchased from Bethesda Research Laboratories, New England Biolabs (Beverly, MA), or Boehringer Mannheim. The pTZ18R plasmid vector and T7 sequencing kit were purchased from Pharmacia LKB Biotechnology Inc. Oligodeoxynucleotides were synthesized and sequence-verified by the oligodeoxynucleotide synthesis service of the University of Pennsylvania Cancer Center. Nucleic acid blot hybridization analyses were performed using Zetabind membrane (AMF Cuno, Meriden, CT) as described previously (61). Southern blots were washed at 68 °C with 75 mM NaCl, 7.5 mM sodium citrate (0.5x SSC), and 0.5% sodium dodecyl sulfate prior to autoradiography. DNA probes were uniformly labeled using a random-priming DNA labeling kit (Boehringer Mannheim). Oligodeoxynucleotides were end-labeled using T4 polynucleotide kinase and [γ-32P]ATP and used for hybridization to Southern blots. The polymerase chain reaction was carried out under conditions that were previously described (62). A λ DASH/EcoRI cloning kit was obtained from Stratagene (La Jolla, CA). All isotopically labeled compounds ([α-32P]deoxynucleotide, 3000 Ci/mmol; [α-32P]dATP, 400 Ci/mmol; and [γ-32P]ATP, 5000 Ci/mmol) were purchased from Amersham Corp. Sequencer gel was obtained from Perkin-Elmer Cetus Instruments.

**Isolation and Characterization of Rat DBP Genomic DNA—**A genomic library generated from Sprague-Dawley rat liver DNA partially digested with EcoRI and then cloned into λ Charon 4A (a gift from T. Sargent, National Institutes of Health) (63) was screened for rDBP genomic clones. The recombinant phage were grown in Escherichia coli DP50SupF. Rat DBP cDNA was used as hybridization probe to identify approximately 105 recombinant phage in the library by in situ hybridization (64). Cloned genomic DNA fragments (clones C1 and C4) were purified and mapped by dot blot hybridization to a series of synthetic oligodeoxynucleotides spanning the entire rDBP cDNA (21; see below). To isolate two missing EcoRI RBB genomic fragments, 10 and 1.4 kb, a size-fractionated EcoRI partial Sprague-Dawley rat liver library was constructed. High molecular weight genomic DNA was partially digested with EcoRI and electrophoresed on a preparative agarose gel, the 10-14 kb size region was excised from the gel, and the DNA was eluted from the gel slice (65). This DNA was phenol-extracted, ethanol-precipitated, and ligated to λ DASH arms pretreated with EcoRI. Recombinant molecules were packaged into phage particles and used to plate using E. coli K804. DNA was phenol-extracted, ethanol-precipitated, and ligated into the EcoRI site of pTZ18R. For one recombinant plasmid (S2, probe S6) sequences were subcloned into the plasmid vector pTZ18R containing rat genomic DNA digested with EcoRI, SstI, Sppl, BamHI, and HindIII.

**DNA Sequence Analysis—**Both single-stranded and double-stranded DNA sequencing were done by the chain termination method (66). Deoxy-7-deaza guanosine (deaza-dGTP) was used in some cases instead of dGTP to alleviate band compressions. Single-stranded DNA templates for sequencing were generated using the pTZ18R recombinant plasmids and M13K07 helper phage in E. coli NM522 host cells. In a few cases, smaller fragments of cloned rDBP genomic DNA were subcloned in M13mp18 and M13mp19 to prepare total RNA-aided DNA. In addition to the universal primer (5' GTAAAACGACGGCCAGT-3') for M13 sequencing and a universal reverse primer (5'-AACAGCTATGACCATG-3') for pTZ18R sequencing, synthetic oligodeoxynucleotides homologous to the rDBP cDNA (21) sense strand beginning at positions 12, 87, 198, 430, 536, 776, 934, 1092, 1218, 1353, and 1513 were also used as primers for sequencing. The oligodeoxynucleotides were chosen to lie within predicted exons and ranged in size from 18 to 20 bp (21). All of the reported rDBP genomic sequence was determined at least twice and in most cases on both strands.

**Primer Extension Analysis—**For primer extension, oligodeoxynucleotides identical to the antisense strand at position 29 (5'- GCCCAAAAGGCTAAGGCGCAGC-3') and at position 102 (5'-GTCCATCCTTTTTCCCCGCTTGA-3') of rDBP cDNA (21) were 5'-end-labeled with [γ-32P]ATP and polynucleotide kinase. 25 µg of rat liver DNA were double-stranded, and DNA was suspended in a small volume of TE (10 mM Tris-HCl, pH 8.0, and 1 mM EDTA) and electrophoresed through a 6% polyacrylamide gel containing 8 M urea next to a DNA sequencing ladder as size marker. The gel was then autoradiographed.

**RESULTS**

**Isolation of Genomic DBP Clones—**λ Charon 4A (63) and λ DASH rat EcoRI genomic libraries were screened with the rDBP cDNA to isolate genomic clones C1 through C4 (Fig. 1A). EcoRI fragments of the genomic clones containing exon sequences were subcloned into the plasmid vector pTZ18R generating subclones S1 through S6 (Fig. 1B). To provide evidence that clones C1 and C2 represent continuous segments of the rDBP gene, rat genomic DNA was digested with SalI, BamHI, and SstI, and these genomic DNA fragments were hybridized with a BamHI/EcoRI fragment from the 3' end of clone C1 and then with an EcoRI/AccI fragment from the 5' end of clone C2 (Fig. 2, IId). Both probes hybridized to DNA fragments of sizes predicted from the cloned fragments: 13.0 kb for SalI, 4.5 kb for BamHI, and 9.75 kb for SstI, indicating that clones C1 and C2 are contiguous (Fig. 2, I and IIa-c). The EcoRI site separating C3 and C4 is internal to rDBP cDNA (Fig. 1A). Polymerase chain reaction of rat...
IR clones hybridizing (stippled box) or not hybridizing (white box) to cDNA in the genomic clones is indicated by the vertical broken line. B, EcoRI fragments of the genomic clones containing coding sequences subcloned in pTZ18R vector are labeled by the restriction maps of the cloned and endogenous DBP gene are identical. Numbered black boxes indicate exons. Restriction sites are abbreviated as follows: B, BamHI; H, HindIII; L, SalI; M, SmaI, R, EcoRI and S, SstI. A polymorphic EcoRI site is indicated (\( \ast \)). D, the direction and extent of DNA sequencing are indicated by the arrows.

Characterization of the rDBP Gene—The location of rDBP mRNA coding sequences in the genomic clones were mapped initially by Southern blot analysis of restriction endonuclease digests of DBP genomic clones hybridized to rDBP cDNA fragments (Fig. 1C). The rDBP exon/intron boundaries were sequenced by the dideoxy chain termination technique, primarily using oligodeoxynucleotide primers derived from the sequence of the rDBP cDNA (position of primers shown in Fig. 1D). DNA sequencing from these sites resulted in the identification of 13 exons. The sequences at the 5' and 3' ends of each intron are in agreement with the consensus sequences for intron/exon boundaries of other eukaryotic genes (Table I) (68). A single copy of the hexanucleotide AATAAA is located 156 bp downstream from the translation termination codon in exon 13 (Fig. 3).

The overall organization of the gene (Fig. 1C) was derived from a combination of the DNA sequence (Fig. 3) and restriction mapping data. The complete restriction map of the gene was generated by hybridization of restriction endonuclease digests to specific oligodeoxynucleotides homologous to different regions of the cDNA and by generating overlapping double digestions of the fragments identified by primary digestions with single enzymes. The final map of the cloned gene was confirmed to be representative of the endogenous rDBP gene, by comparing it with Southern blots of total genomic DNA digested with the same enzymes (data not shown). A polymorphic EcoRI site was identified in intron 6 of the cloned DBP gene (Fig. 1C) which was absent in the rat used for total genomic DNA mapping. The rDBP gene is approximately 35 kb from the cap site to the polyadenylation site. The first intron is the largest, containing 10.45 kb. Three small introns, 2, 5, and 9, were completely sequenced. Genomic clone C1 includes 2196 bp of 5'-flanking region which was sequenced in entirety, and clone C4 contains about 2.4 kb of 3'-flanking region.

The coding sequence in the Sprague-Dawley rDBP gene is identical to the Sprague-Dawley kidney rDBP cDNA sequence (62) except at two positions. The first base of the codon for the 116th amino acid is G in the gene but C in the kidney cDNA (62), resulting in a Glu in the gene instead of the Gln reported in the cDNA. However, the liver rDBP cDNA (21) contains a codon identical to that found in gene at this position. The 28th base upstream from the polyadenylation signal (AATAAA) is a T in the gene and liver rDBP cDNA but a G in the kidney rDBP cDNA (62). No other discrepancies were detected.

To determine DBP gene copy number, restriction digests of total rat genomic DNA were analyzed by Southern blotting using 5'- and 3'-flanking region probes. The probes from each end of the gene hybridized to a single band which corresponded to the size predicted from the cloned gene (data not shown). Based on these experiments, we predict that the rDBP locus exists as a single copy in the haploid rat genome.

Repetitive Sequences—By computer-generated dot matrix analysis comparing the 5' regions of the hDBP and rDBP genes, two discrete rDBP domains were noted to share 52–58% sequence identity with the similar region of the hDBP gene flanking region (Fig. 3, underlined). In the hDBP gene,
these two segments are adjacent, while in the rDBP gene they are separated by 607 bp containing within it a 180-bp segment with 87% sequence identity to the consensus sequence for the murine B2 family of repetitive elements (69) and 86% identity to each of two similar middle repetitive elements associated with the rat growth hormone gene (69, 70). This repetitive element in the rDBP 5'-flanking region contains the consensus sequences for the split polymerase III promoter and the oligo(dT) stretches characteristic of polymerase III termination (71) (Fig. 3). These elements suggest that this repetitive sequence has a transcriptional orientation opposite that of the adjacent DBP gene. The activity of this putative transcription unit is not known. The repetitive element is flanked by 25- and 23-bp polypyrimidine tracts which are T-rich, nearly identical, inverted repeats.

Extensive homopurine polymer tracts were identified in introns 1, 2, and 5. Tracts of alternating G and T were separated by 607 bp containing within it a 180-bp segment with 87% sequence identity to the consensus sequence for the murine B2 family of repetitive elements (69) and 86% identity to each of two similar middle repetitive elements associated with the rat growth hormone gene (69, 70). This repetitive element in the rDBP 5'-flanking region contains the consensus sequences for the split polymerase III promoter and the oligo(dT) stretches characteristic of polymerase III termination (71) (Fig. 3). These elements suggest that this repetitive sequence has a transcriptional orientation opposite that of the adjacent DBP gene. The activity of this putative transcription unit is not known. The repetitive element is flanked by 25- and 23-bp polypyrimidine tracts which are T-rich, nearly identical, inverted repeats.

Extensive homopurine polymer tracts were identified in introns 1, 2, and 5. Tracts of alternating G and T were separated by 607 bp containing within it a 180-bp segment with 87% sequence identity to the consensus sequence for the murine B2 family of repetitive elements (69) and 86% identity to each of two similar middle repetitive elements associated with the rat growth hormone gene (69, 70). This repetitive element in the rDBP 5'-flanking region contains the consensus sequences for the split polymerase III promoter and the oligo(dT) stretches characteristic of polymerase III termination (71) (Fig. 3). These elements suggest that this repetitive sequence has a transcriptional orientation opposite that of the adjacent DBP gene. The activity of this putative transcription unit is not known. The repetitive element is flanked by 25- and 23-bp polypyrimidine tracts which are T-rich, nearly identical, inverted repeats.

Extensive homopurine polymer tracts were identified in introns 1, 2, and 5. Tracts of alternating G and T were separated by 607 bp containing within it a 180-bp segment with 87% sequence identity to the consensus sequence for the murine B2 family of repetitive elements (69) and 86% identity to each of two similar middle repetitive elements associated with the rat growth hormone gene (69, 70). This repetitive element in the rDBP 5'-flanking region contains the consensus sequences for the split polymerase III promoter and the oligo(dT) stretches characteristic of polymerase III termination (71) (Fig. 3). These elements suggest that this repetitive sequence has a transcriptional orientation opposite that of the adjacent DBP gene. The activity of this putative transcription unit is not known. The repetitive element is flanked by 25- and 23-bp polypyrimidine tracts which are T-rich, nearly identical, inverted repeats.

Extensive homopurine polymer tracts were identified in introns 1, 2, and 5. Tracts of alternating G and T were separated by 607 bp containing within it a 180-bp segment with 87% sequence identity to the consensus sequence for the murine B2 family of repetitive elements (69) and 86% identity to each of two similar middle repetitive elements associated with the rat growth hormone gene (69, 70). This repetitive element in the rDBP 5'-flanking region contains the consensus sequences for the split polymerase III promoter and the oligo(dT) stretches characteristic of polymerase III termination (71) (Fig. 3). These elements suggest that this repetitive sequence has a transcriptional orientation opposite that of the adjacent DBP gene. The activity of this putative transcription unit is not known. The repetitive element is flanked by 25- and 23-bp polypyrimidine tracts which are T-rich, nearly identical, inverted repeats.

Extensive homopurine polymer tracts were identified in introns 1, 2, and 5. Tracts of alternating G and T were separated by 607 bp containing within it a 180-bp segment with 87% sequence identity to the consensus sequence for the murine B2 family of repetitive elements (69) and 86% identity to each of two similar middle repetitive elements associated with the rat growth hormone gene (69, 70). This repetitive element in the rDBP 5'-flanking region contains the consensus sequences for the split polymerase III promoter and the oligo(dT) stretches characteristic of polymerase III termination (71) (Fig. 3). These elements suggest that this repetitive sequence has a transcriptional orientation opposite that of the adjacent DBP gene. The activity of this putative transcription unit is not known. The repetitive element is flanked by 25- and 23-bp polypyrimidine tracts which are T-rich, nearly identical, inverted repeats.

Extensive homopurine polymer tracts were identified in introns 1, 2, and 5. Tracts of alternating G and T were separated by 607 bp containing within it a 180-bp segment with 87% sequence identity to the consensus sequence for the murine B2 family of repetitive elements (69) and 86% identity to each of two similar middle repetitive elements associated with the rat growth hormone gene (69, 70). This repetitive element in the rDBP 5'-flanking region contains the consensus sequences for the split polymerase III promoter and the oligo(dT) stretches characteristic of polymerase III termination (71) (Fig. 3). These elements suggest that this repetitive sequence has a transcriptional orientation opposite that of the adjacent DBP gene. The activity of this putative transcription unit is not known. The repetitive element is flanked by 25- and 23-bp polypyrimidine tracts which are T-rich, nearly identical, inverted repeats.

Extensive homopurine polymer tracts were identified in introns 1, 2, and 5. Tracts of alternating G and T were separated by 607 bp containing within it a 180-bp segment with 87% sequence identity to the consensus sequence for the murine B2 family of repetitive elements (69) and 86% identity to each of two similar middle repetitive elements associated with the rat growth hormone gene (69, 70). This repetitive element in the rDBP 5'-flanking region contains the consensus sequences for the split polymerase III promoter and the oligo(dT) stretches characteristic of polymerase III termination (71) (Fig. 3). These elements suggest that this repetitive sequence has a transcriptional orientation opposite that of the adjacent DBP gene. The activity of this putative transcription unit is not known. The repetitive element is flanked by 25- and 23-bp polypyrimidine tracts which are T-rich, nearly identical, inverted repeats.

Extensive homopurine polymer tracts were identified in introns 1, 2, and 5. Tracts of alternating G and T were separated by 607 bp containing within it a 180-bp segment with 87% sequence identity to the consensus sequence for the murine B2 family of repetitive elements (69) and 86% identity to each of two similar middle repetitive elements associated with the rat growth hormone gene (69, 70). This repetitive element in the rDBP 5'-flanking region contains the consensus sequences for the split polymerase III promoter and the oligo(dT) stretches characteristic of polymerase III termination (71) (Fig. 3). These elements suggest that this repetitive sequence has a transcriptional orientation opposite that of the adjacent DBP gene. The activity of this putative transcription unit is not known. The repetitive element is flanked by 25- and 23-bp polypyrimidine tracts which are T-rich, nearly identical, inverted repeats.

Extensive homopurine polymer tracts were identified in introns 1, 2, and 5. Tracts of alternating G and T were separated by 607 bp containing within it a 180-bp segment with 87% sequence identity to the consensus sequence for the murine B2 family of repetitive elements (69) and 86% identity to each of two similar middle repetitive elements associated with the rat growth hormone gene (69, 70). This repetitive element in the rDBP 5'-flanking region contains the consensus sequences for the split polymerase III promoter and the oligo(dT) stretches characteristic of polymerase III termination (71) (Fig. 3). These elements suggest that this repetitive sequence has a transcriptional orientation opposite that of the adjacent DBP gene. The activity of this putative transcription unit is not known. The repetitive element is flanked by 25- and 23-bp polypyrimidine tracts which are T-rich, nearly identical, inverted repeats.

Extensive homopurine polymer tracts were identified in introns 1, 2, and 5. Tracts of alternating G and T were separated by 607 bp containing within it a 180-bp segment with 87% sequence identity to the consensus sequence for the murine B2 family of repetitive elements (69) and 86% identity to each of two similar middle repetitive elements associated with the rat growth hormone gene (69, 70). This repetitive element in the rDBP 5'-flanking region contains the consensus sequences for the split polymerase III promoter and the oligo(dT) stretches characteristic of polymerase III termination (71) (Fig. 3). These elements suggest that this repetitive sequence has a transcriptional orientation opposite that of the adjacent DBP gene. The activity of this putative transcription unit is not known. The repetitive element is flanked by 25- and 23-bp polypyrimidine tracts which are T-rich, nearly identical, inverted repeats.

Extensive homopurine polymer tracts were identified in introns 1, 2, and 5. Tracts of alternating G and T were separated by 607 bp containing within it a 180-bp segment with 87% sequence identity to the consensus sequence for the murine B2 family of repetitive elements (69) and 86% identity to each of two similar middle repetitive elements associated with the rat growth hormone gene (69, 70). This repetitive element in the rDBP 5'-flanking region contains the consensus sequences for the split polymerase III promoter and the oligo(dT) stretches characteristic of polymerase III termination (71) (Fig. 3). These elements suggest that this repetitive sequence has a transcriptional orientation opposite that of the adjacent DBP gene. The activity of this putative transcription unit is not known. The repetitive element is flanked by 25- and 23-bp polypyrimidine tracts which are T-rich, nearly identical, inverted repeats.

Extensive homopurine polymer tracts were identified in introns 1, 2, and 5. Tracts of alternating G and T were separated by 607 bp containing within it a 180-bp segment with 87% sequence identity to the consensus sequence for the murine B2 family of repetitive elements (69) and 86% identity to each of two similar middle repetitive elements associated with the rat growth hormone gene (69, 70). This repetitive element in the rDBP 5'-flanking region contains the consensus sequences for the split polymerase III promoter and the oligo(dT) stretches characteristic of polymerase III termination (71) (Fig. 3). These elements suggest that this repetitive sequence has a transcriptional orientation opposite that of the adjacent DBP gene. The activity of this putative transcription unit is not known. The repetitive element is flanked by 25- and 23-bp polypyrimidine tracts which are T-rich, nearly identical, inverted repeats.
The nucleotide sequence of the rat DBP gene and its flanking regions. The numbers to the right of each line correspond to the last nucleotide in that line, and the numbers in the middle of the line above the triplet codons represent amino acid numbers. The cap site is assigned *1*. Exons are capitalized, and intron and flanking sequences are in lower case. Nonsequenced regions of introns are indicated by dots. Estimated sizes of each exon and intron are in parentheses. **Fig. 3**—continued.
### DBP Gene Structure

#### Table II

**Similarity between the proximal and distal regulatory elements of ALB, AFP, and DBP genes**

Only those putative regulatory elements shared between ALB, AFP, and DBP genes are included. Identified regulatory elements are indicated by boldface; the remainder are putative. The putative elements are assigned on the basis of sequence similarity and location of the sequence motif. Consensus GREs are underlined. Numbers at the top of each sequence represent the location of that sequence motif relative to transcription initiation site. NF, not found; ND, not determined.

| Gene | AFP enhancer core | DEII | GRE | HNF-1 | CAAT | TATA | Refs. |
|------|-------------------|------|-----|-------|------|------|-------|
| ALB  |                   |      |     |       |      |      |       |
| Human| NF                | -129 | -177| -65   | -88-84| -32-26| 41, 49, 27 |
|      |                   | TTTTGGCAAGGA |      | TAGTTAATAATCAGAA | CCAAT | TATATTA |
|      |                   | -358 | -342| -50-84| -33-27|       | 42, 50 |
| Mouse| NF                | -129 | -117| -60   | -88-84| -33-27| 43, 51 |
|      |                   | TTTTGGCAAGGA |      | TAGTTAATGATGACTACAG | CCAAT | TATATTA |
|      |                   | -67   | -51 | -89-85| -34-28|       | 52, 28 |
| Rat  | NF                | -130 | -118| -67   | -89-85| -34-28| 52, 28 |
|      |                   | TTGTTGCAAGGA |      | TAGTTAATGATGACTACAG | CCAAT | TATATTA |
|      |                   | -67   | -51 | -89-85| -34-28|       | 52, 28 |
| Xenopus| NF              |      |     |       |      |      |       |
|      |                   |      |     | AGGTTAATATTTC | NF | TATAAA | 53, 45 |
|      |                   | AFAGCCTCAGGTTG | | AGGTTACTAATGATA | CCAAC | TATATTA |
|      |                   | -132 | -116|       |      |      | 48, 30 |
|      |                   | CTGTTAATTATGGGCAA | |       |      |      | 54 |
| Human|                   | -8500|      |       |      |      | 54 |
|      |                   | TATATATATTACAA | |       |      |      | 54 |
| Mouse|                   | -159 | -159| -64   | -90-24|       | 40, 23 |
|      |                   | AGTTGCTTTTGGCAAGG | | AGGTTACTAATGACTACAG | NF | TATAAA |
|      |                   | -130 | -114|       |      |      | 40, 55 |
|      |                   | CTGTTAATATTGGGCAA | |       |      |      | 40, 55 |
| Mouse|                   | -59   |      |       |      |      | 40 |
|      |                   | TTTTGGCTACA | |       |      |      | 40 |
| Rat  | ND                | -172 | -158| -63   | -70-66| -29-23| 58, 65, 33 |
|      |                   | AGTTGCTTTTGGCAAGG | | AGGTTACTAATGACTACAG | NF | TATATTA |
|      |                   | -1900 | -113|       |      |      | 58, 65, 33 |
|      |                   | CTGTTAATATTGGGCAA | |       |      |      | 48, 57 |
| DBP  |                   |      |     |       |      |      |       |
| Human|                   | -127 | -114| -65   | -89-84| -34-28| 52, 28 |
|      |                   | TTTTGGGCAAGGA | | AGATTAATATTGATTA | NF | TATATTA |
|      |                   | -166-1689|      |       |      |      | 59 |
|      |                   | GGGTTTACTG | |       |      |      | 59 |
| Rat  |                   | -152 | -152| -61   | -91-87| -25-19| 52, 70 |
|      |                   | CTCCGACTGCGCCTT | | AGATTAATATTGATTA | CCAAT | TATATTA |
|      |                   | -183 | -167|       |      |      | 52, 70 |
|      |                   | AGGTTAATATTGATAAAC | |       |      |      | 52, 70 |
| Rat  |                   | -1689 | -1689|       |      |      | 52, 70 |
|      |                   | GTCATACTGCTT | |       |      |      | 52, 70 |
|      |                   | -2044 | -2030|       |      |      | 52, 70 |
|      |                   | -2060 | -2044|       |      |      | 52, 70 |
|      |                   | -2060 | -2044|       |      |      | 52, 70 |

**Consensus sequences**

| TGGTGCAAGGA | TGGTTGCAAGGA | TGGTTGCAAGGA | TGGTTGCAAGGA | TGGTTGCAAGGA | TGGTTGCAAGGA |
|-------------|-------------|-------------|-------------|-------------|-------------|
| ACCCTCCTCTT | ACCCTCCTCTT | ACCCTCCTCTT | ACCCTCCTCTT | ACCCTCCTCTT | ACCCTCCTCTT |
| TGGTTAATATT | TGGTTAATATT | TGGTTAATATT | TGGTTAATATT | TGGTTAATATT | TGGTTAATATT |

exon 15 contains only untranslated sequences, comparable to exons 12 and 13 of rDBP. Previous comparisons of the encoded proteins noted that DBP is shorter by 124 residues, reflecting a truncation in the third internal repeat motif. The most direct mechanism for this would have been the imposition of a nonsense mutation in the corresponding region of the gene, that is in exon 12. From the data in the present report, it becomes clear that this is not the case. Instead, as initially suggested by Gibbs and Dugaiczyk (35), it appears that truncation of DBP most likely occurred by loss of two internal exons from the DBP progenitor gene after its divergence from the ALB and AFP lineage (Fig. 4). Progenitor exons 12 and 13 could have been lost from the DBP gene either by deletion or by a nondeleterious defect such as a splicing site mutation. It is not possible to propose a mechanism based on the available data. It is noteworthy, however, that intron 11 of the rDBP which might be expected to contain remnants of the progenitor exons (Fig. 4) is the second largest intron in the DBP gene (4.31 kb) and spans a distance of approximately the same length as the distance from exons 11 to 14 in the AFP and ALB genes. However, our attempts to detect these sequences in intron 11 by low stringency hybridization to ALB and AFP probes were unsuccessful (data not shown). This suggests that if the loss was by a nondeleterious event, it must have occurred in the remote past to have allowed for such a substantial drift in sequence.

The structural characterization of the regions flanking the DBP gene shed further light on the evolution of the DBP/ALB/AFP gene family. Southern analysis of the 5'- and 3'-flanking regions of the DBP gene reveal restriction patterns identical to those found in the cloned gene. This identity strongly suggests that there is only a single copy of the DBP gene in the haploid rat genome. Comparison of the 5'-flanking regions of the hDBP and rDBP genes at the level of primary structure reveals that the rat gene contains two discrete regions of sequence similarity to hDBP: (a) positions -1957 to -1203 with 58% identity and (b) positions -595 to -1 with 52% identity. A segment of the nucleotide sequence between these two regions contains 87% identity to the middle repetitive elements mouse BZ (69) and rat RU (69, 70). This segment is flanked by an imperfect inverted polypyrimidine repeat (Fig. 3), suggesting that it was inserted adjacent to the
rat gene by a transposon-mediated event. Since this element contains a segment homologous to the split RNA polymerase III promoter, it is conceivable that it may be transcriptionally active (69). However, since it is oriented on the minus strand, these transcripts would not be expected to interfere with the rDBP gene proximal promoter region as has been postulated for a similar element found internal to the first intron of the mouse AFP gene (71, 77). The presence of this repetitive element in rat but not human suggests that it was inserted 5' of the DBP locus subsequent to the divergence of the human and rat lineage approximately 85 million years ago (78).

Despite strong identity in nucleotide and amino acid sequence and a clear evolutionary relationship among the DBP, ALB, and AFP genes, it was initially reported on the basis of a partial characterization of the hDBP 5'-flanking region that these genes were dissimilar in promoter and proximal promoter elements (59). Based upon our complete sequence analysis of this region (Fig. 3) and a detailed comparison to the corresponding regions of ALB and AFP (Table II), we find that these genes do in fact have a number of evolutionarily preserved sequence motifs. The rDBP gene may have two transcriptional start sites. The major start site is located 62 bp upstream of the translation initiation (AUG) codon, with a possible minor start site located 11 nucleotides further downstream. The transcriptional start site of the hDBP gene has been reported to be similarly positioned at 61 bases upstream from its AUG codon (59) with a second weaker putative site located 92 bp further upstream (22). Both the human and rat DBP genes contain TGTAAA at approximately -25 to their major transcription start sites (Table II). In a nonexhaustive study, a wide variety of DNA sequences have been shown to functionally replace a yeast TATA element including a strongly active CATTTAAT and TGTGCG (79), supporting the possibility that TGTAAA may be active. Unlike ALB (80), neither AFP nor DBP contain a classical CCAAT box. However, the rDBP gene contains the pentamer CAAAT at position -91 at the same distance from the promoter as the functionally active CCAAT in ALB (Table II). Therefore, it appears feasible that TGTAAA serves the role of the canonical TATA for the DBP gene transcription, as suggested in the case of the low density lipoprotein receptor gene (81), and that the CAAAT may substitute for the canonical CCAAT sequence located further upstream.

The basis for tissue-specific and developmentally specific expression of the ALB/DBP/DBP genes is of central interest. All the members of this multigene family are expressed in liver. The DNA binding protein HNF-1 is thought to be responsible for the liver-specific expression of a number of genes including ALB and AFP (82). An HNF-1 cis element has been identified at similar positions in the human (54), mouse (55), and rat (56, 57) AFP promoters. More remarkably, the position and sequence of the HNF-1 cis element in the 5'-flanking region of the ALB genes has been conserved from Xenopus to humans (82). A putative HNF-1 binding site is present in the proximal promoters of both human and rat DBP genes in positions similar to its location in the ALB and AFP genes (around -65) (Table II). AFP and DBP, but not ALB, have an additional HNF-1 consensus region further upstream between bases -110 and -210. The conservation of the HNF-1 site in all three of these genes is consistent with their high levels of expression in the liver.

The DBP/ALB/FP genes share transcriptional responses to a number of environmental stimuli. These shared responses may be mediated by a series of conserved proximal promoter elements. The AFP gene is negatively regulated by dexameth-

---

**Table III**

Sizes (in nucleotides) of exons of ALB/FP/DBP family genes

The numbers in parentheses for exons 1, 14, and 15 represent the contributions from the coding sequence including the termination codons. Sequences for human ALB (27) and AFP (30), mouse AFP (23, 34), and rat ALB (28, 29) and AFP (31, 32) were obtained from the published literature. Some exon sizes for rat AFP (square brackets) were assigned on the basis of extensive sequence similarity to the respective mouse sequence (35).

| Exon | Human | Mouse | Rat | DBP |
|------|-------|-------|-----|-----|
| 1    | 118 (79) | 129 (85) | 129 (85) | 117 (79) | [126 (103)] | 129 (58) |
| 2    | 58    | 52    | 52    | 58    | 52    | 70    |
| 3    | 133   | 133   | 121   | 133   | [121] | 133   |
| 4    | 212   | 212   | 212   | 212   | [212] | 212   |
| 5    | 133   | 133   | 133   | 133   | 133   | 133   |
| 6    | 98    | 98    | 98    | 98    | 98    | 95    |
| 7    | 130   | 130   | 130   | 130   | 130   | 130   |
| 8    | 215   | 215   | 215   | 215   | 215   | 203   |
| 9    | 133   | 133   | 133   | 133   | 133   | 130   |
| 10   | 98    | 98    | 98    | 98    | [98] | 98    |
| 11   | 139   | 139   | 139   | 139   | [139] | 133   |
| 12   | 224   | 224   | 224   | 224   | [224] |       |
| 13   | 133   | 133   | 133   | 133   | [133] |       |
| 14   | 68 (45) | 55 (45) | 55 (45) | 62 (42) | 55 (45) | 60 (36) |
| 15   | 163 (0) | 145 (0) | 141 (0) | 140 (0) | [141 (0)] | 159 (0) |
asone during intrauterine life (74, 75). Consistent with this, the human and rat AFP genes contain functional GREs (48, 58) at about -170, and a putative GRE is present in mouse AFP (Table II). The rDBP gene whose regulation by glucocorticoid has not been rigorously studied contains a putative GRE with core element sequence TGTCCT at position -157. Of note, the hDBP gene contains a putative GRE with the variant core consensus TGTCCTA at -161 (59). The ALB gene is not known to be regulated by glucocorticoid and lacks a GRE. DBP has two additional reported responses to steroid hormones, an increase in rat serum DBP levels in response to androgens (83) and an increase in human serum DBP levels in response to estrogens (84). Neither can be correlated with specific proximal promoter motifs at this point. Future experiments on regulation of the DBP gene at the molecular level should clarify the molecular basis for these observations. A number of additional cis-acting transcriptional motifs of less well defined function can be identified in the 5'-flanking region of DBP and either ALB or AFP. Based upon the sequence data generated in the present report, we have identified and summarized these in Table II. These include the DEFI motif at position -127 in human but not rat DBP. This element mediates transcriptional activation of ALB by binding to the nuclear transcription factor NF1 (41). In the AFP promoter region, a few additional motifs were detected. The AFP enhancer core element (TGTTCGA/TGT) located at -5.0 kb, -2.0 kb, and -78 bp of the AFP gene (40) was detected 5' to the rDBP gene at position -1698 as it has been in other liver-specific genes (85-87). An HNF-1-like protein in hepatoma cells binds to an HNF-1 motif at -3.5 kb in the AFP gene (54), and similarly the rDBP gene has an HNF-1 consensus site at a distal promoter position, -2060. hDBP sequence data does not extend far enough to determine its presence or absence. The rAFP gene has also been shown to bind the glucocorticoid receptor at a distal location around position -1.9 kb (48), and by analogy, we find that the rDBP gene contains a distal promoter GRE consensus sequence (TGTTCCT) at position -2035. The developmental patterns of ALB and DBP are quite similar and are clearly distinct from that of AFP. Both DBP and ALB are expressed at high levels in adult liver, while the AFP gene is expressed at high levels in fetal yolk sac and regenerating liver but not in the normal adult liver. Unexpectedly, our search for conserved features revealed that the DBP promoter region resembles the AFP promoter more closely than that of ALB. Both DBP and AFP lack a canonical CCAAT motif, contain an additional HNF-1 binding site at similar locations in the proximal and distal promoter regions, contain GREs in similar proximal and distal locations, and have a similar enhancer core element in distal locations. The biological significance of the sequence motifs identified in DBP based upon sequence similarity and evolutionary conservation arguments must remain speculative until their roles in transcription are defined by functional assays. It is now possible to compare the overall genomic structures of the three members of this multigene family (Fig. 4). The DBP gene, previously postulated to be the oldest member of this family (21), is the largest, consistent with the finding in other gene families that the oldest member has accumulated the largest amount of intronic DNA (88). Despite its large size, DBP has two fewer exons than ALB and AFP, reflecting the probable loss of primordial exons 12 and 13 during evolution. The divergence of exons 1 and 2 and the loss of exons 12 and 13 resulted in divergence of protein functions such as the ability of DBP to bind 25(OH)D3 and the loss of a high affinity binding site for unsaturated fatty acids as compared with ALB and AFP. Comparisons of the promoter region are remarkable for the conservation of regulatory features, many of which have been confirmed biologically for the ALB and AFP genes and are currently under study for the DBP gene.

Acknowledgments—The authors thank Kai Voigt and Lara Soowal for their contributions to this work, Dr. Stephen Liebhaber for many helpful suggestions, and Wil Meredith for preparing the manuscript.

REFERENCES

1. Hay, A. W. M., and Watson, G. (1976) Comp. Biochem. Physiol. 53B, 167-172
2. Haddad, J. G., Jr., and Walgate, J. (1976) J. Biol. Chem. 251, 4803-4809
3. Bouillon, R., Van Baelen, H., Rombouts, W., and de Moor, P. (1978) J. Biol. Chem. 253, 4425-4431
4. Smithies, O. (1985) Biochem. J. 211, 629-641
5. Haddad, J. G. (1984) in Vitamin D (Kumar, R., ed) pp. 383-395, Martinus Nijhoff, Boston
6. Haddad, J. G., Hillman, L., and Roojasanathit, S. (1976) J. Clin. Endocrinol. & Metab. 43, 86-91
7. McLeod, J. F., Kowalski, M. A., and Haddad, J. G., Jr. (1989) J. Biol. Chem. 264, 1250-1257
8. Harper, K. D., McLeod, J. F., Kowalski, M. A., and Haddad, J. G. (1987) J. Clin. Invest. 79, 1365-1370
9. Lind, S. E., Smith, D. B., Janney, P. A., and Stossel, T. P. (1986) J. Clin. Invest. 78, 736-742
10. Haddad, J. G., Harper, K. D., Gooth, M., Pietra, G. G., and Singer, J. W. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 1381-1385
11. Sanger, J. M., Dabiri, G., Mittal, B., Kowalski, M. A., Haddad, J. G., and Sanger, J. W. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 5474-5478
12. Kew, R. R., and Webster, R. O. (1988) J. Clin. Invest. 82, 364-369
13. Petrini, M., Galbraith, R. M., Emerson, D. L., Nel, A. E., and Arnaud, P. (1985) J. Biol. Chem. 260, 1804-1810
14. Petrini, M., Emerson, D. L., and Galbraith, R. M. (1983) Nature 306, 73-76
15. McLeod, J. F., Kowalski, M. A., and Haddad, J. G. (1986) Endocrinology 119, 77-83
16. Nestler, J. E., McLeod, J. F., Kowalski, M. A. Strauss, J. F., III, and Haddad, J. G., Jr. (1987) Endocrinology 120, 1996-2002
17. Gooth, M., Murgia, A., Smith, R. M., Prystowsky, M. B., Cooke, N. E., and Haddad, J. G. (1990) Endocrinology 127, 2313-2321
18. Mikkelson, M., Jacobsen, P., and Henningsen, K. (1977) Hum. Hered. 27, 100-102
19. Cleve, H. (1973) J. Med. Sci. 9, 1133-1146
20. Cooke, N. E., and David, E. V. (1985) J. Clin. Invest. 76, 2420-2424
21. Cooke, N. E. (1986) J. Biol. Chem. 261, 3441-3450
22. Yang, F., Brune, J. L., Naylor, S. L., Cupples, R. L., Nabehaus, K. H., and Bowman, B. H. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 7994-7998
23. Eiferman, F. A., Young, P. R., Scott, R. W., and Tilghman, S. M. (1981) Nature 294, 713-718
24. Brown, J. R. (1976) Fed. Proc. 35, 2141-2144
25. Cooke, N. E., Willard, H. G., David, E. V., and George, D. L. (1986) Hum. Genet. 73, 225-229
26. Cooke, N. E., Levan, G., and Sapir, J. (1987) Cytogenet. Cell Genet. 44, 96-100
27. Minnghetti, P. P., Ruffner, D. E., Kuang, W.-J., Hawkins, J. W., Beattie, W. G., and Dugaiczyk, A. (1988) J. Biol. Chem. 261, 6747-6757
28. Sargent, T. D., Jagodzinski, L. L., Yang, M., and Bonner, J. (1981) Mol. Cell. Biol. 1, 871-883
29. Sargent, T. D., Yang, M., and Bonner, J. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 243-246
30. Gibbs, P. E. M., Zielinski, R., Boyd, C., and Dugaiczyk, A. (1987) Biochemistry 26, 1322-1343
31. Terrotte, B., Guerret, M., Chevrett, M., and Blanger, L. (1985) Nucleic Acids Res. 13, 2875-2898
32. Jagodzinski, L. L., Sargent, T. D., Yang, M., Giackin, C., and Bonner, J. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 3521-3525
33. Nahon, J.-L., Danan, J.-L., Poiré, M., Tratner, I., Jose-Estanyol, L., and Haddad, J. G., Jr. (1982) J. Clin. Endocrinol. & Metab. 6228
34. Acknowledgments—The authors thank Kai Voigt and Lara Soowal for their contributions to this work, Dr. Stephen Liebhaber for many helpful suggestions, and Wil Meredith for preparing the manuscript.
