

Structure and nucleotide sequence of the rat intestinal vitamin D-dependent calcium binding protein gene

JOHANN KRISINGER*, HISHAM DARWISH*, NOBOYA MAEDA†, AND HECTOR F. DELUCA*‡

Departments of *Biochemistry and †Genetics, University of Wisconsin–Madison, Madison, WI 53706

Contributed by Hector F. DeLuca, September 6, 1988

ABSTRACT The vitamin D-dependent intestinal calcium binding protein (ICaBP, 9 kDa) is under transcriptional regulation by 1,25-dihydroxyvitamin D₃ [1,25-(OH)₂D₃], the hormonal active form of the vitamin. To study the mechanism of gene regulation by 1,25-(OH)₂D₃, we isolated the rat ICaBP gene by using a cDNA probe. Its nucleotide sequence revealed 3 exons separated by 2 introns within ≈3 kilobases. The first exon represents only noncoding sequences, while the second and third encode the two calcium binding domains of the protein. The gene contains a 15-base-pair imperfect palindrome in the first intron that shows high homology to the estrogen-responsive element. This sequence may represent the vitamin D-responsive element involved in the regulation of the ICaBP gene. The second intron shows an 84-base-pair-long simple nucleotide repeat that implicates Z-DNA formation. Genomic Southern analysis shows that the rat gene is represented as a single copy.

The vitamin D-dependent intestinal calcium binding protein (ICaBP, 9 kDa) belongs to the family of intracellular calcium binding proteins involved in regulating calcium metabolism (1). In addition to its presence in the duodenum (2–4), it has been detected in other tissues such as uterus (5) and placenta (6, 7). The protein has two calcium binding sites with each folded in the indicated EF hand structure (1, 8) shared by other calcium binding proteins such as calmodulin and troponin C (1). The amino acid sequences of porcine (9) and bovine (10) calcium binding proteins have been determined directly from the protein. Furthermore, a partial rat protein sequence has been derived from mRNA (11) or from *in vitro* translation (12).

Another 28-kDa vitamin D-dependent calcium binding protein found in chicken intestine has four binding sites for calcium and binds 4–6 mol of Ca per mol of protein (13, 14). In mammals, the 28-kDa protein is not found in intestine but is expressed in kidney (15) and brain (16). The amino acid sequence of this protein has also been determined from the mRNA (17) or directly from the protein (18). A number of other proteins have been reported to be modulated by 1,25-dihydroxyvitamin D₃ [1,25-(OH)₂D₃] at the level of transcription. This includes the parathyroid hormone (19), bone collagen type I (20), *c-myc* gene (21), metallothionein (22), and osteopontin (23). Nevertheless, the 9-kDa vitamin D-dependent ICaBP is the best characterized of the proteins for its dependence on vitamin D at the molecular level (24–26). Although it is believed that 1,25-(OH)₂D₃ activates transcription of the ICaBP gene through an interaction of the ligand–receptor complex with some specific regulatory elements in the gene in accordance with the other steroid hormones' mechanism of action (27), no detailed steps involved in this process for 1,25-(OH)₂D₃ have yet been described. To achieve such an objective, the isolation of the

ICaBP gene as a model to study the molecular mechanism of vitamin D action is an obvious task.

We have previously cloned the cDNA of the ICaBP from an intestinal rat cDNA library (28). The cDNA has 406 nucleotides (nt) with an open reading frame of 237 nt encoding 79 amino acids. In addition, a genomic clone containing the gene for this protein was isolated (28). In the present paper, we report the complete nucleotide sequence of the gene and some of its structural features.[§] By analogy we can suggest a possible regulatory sequence element within the gene that is likely the site of binding of the receptor–ligand complex.

MATERIALS AND METHODS

Isolation and Characterization of the Rat ICaBP Gene. A genomic recombinant clone comprising the entire rat ICaBP was isolated from a partial *EcoRI* genomic library in λ Charon 4 and characterized as described (28).

DNA Sequence Analysis. The nucleotide sequence of the isolated clone was established by the chemical degradation methods of Maxam and Gilbert (29). Sequence information was obtained from both strands of three sets of restriction endonuclease digests that covered the entire genomic clone. Computer analysis of the DNA sequence was carried out using the software provided by the genetics computer group of the University of Wisconsin–Madison (30).

Primer Extension. Poly(A)⁺ RNA from the intestine of animals maintained on a low calcium and vitamin D-supplemented diet (33) was isolated by the guanidine hydrochloride method (31) followed by two steps of oligo(dT) chromatography (32). Two 20-mer oligonucleotides made complementary to the cDNA sequence at positions 20–39 and 387–406 were used as primers. After 5'-end-labeling with polynucleotide kinase and [γ -³²P]ATP (5000 Ci/mmol; 1 Ci = 37 GBq; Amersham) 1.6 ng of primer was hybridized to 25 μ g of RNA at 42°C for 10 min. Ten units of avian myeloblastosis virus reverse transcriptase (Life Sciences, Saint Petersburg, FL) and 4 nmol of dNTP mixture were added and incubated for another 30 min at 42°C. The DNA was ethanol-precipitated and run on a 6% polyacrylamide sequencing gel. Size markers were generated from plasmid pBR322 DNA digested with *Msp* I and filled in with [α -³²P]dCTP (3000 Ci/mmol; Amersham) by DNA polymerase I Klenow fragment. The gel was dried and exposed to Kodak X-Omat film (Eastman Kodak).

Genomic DNA Preparation. Total genomic DNA for Southern analysis was isolated from male rats (Sprague–Dawley). Rats were killed by decapitation and the duodenum, liver, and kidney were removed and rinsed in ice-cold saline. The mucosa was harvested from the duodenum by scraping.

Abbreviations: ICaBP, intestinal calcium binding protein; 1,25-(OH)₂D₃, 1,25-dihydroxyvitamin D₃; nt, nucleotide(s).

[‡]To whom correspondence should be addressed at: Department of Biochemistry, University of Wisconsin–Madison, 420 Henry Mall, Madison, WI 53706. (No reprints will be available from the authors.)

[§]The sequence reported in this paper is being deposited in the EMBL/GenBank data base (accession no. J04133).

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

Genomic DNA was prepared from all tissues according to Petropoulos *et al.* (34) with some modifications. The DNA was extracted with phenol/chloroform and precipitated. The final pellet was dissolved in 10 nM sodium acetate (pH 5.6), sheared, and stored at -70°C .

DNA Digestion and Southern Analysis. DNA (10 μg) was digested with an excess of restriction endonuclease (*Hind*III, 4 units/ μg) at 37°C for 10–12 hr and fractionated on 0.8% agarose. The DNA was transferred to nylon membranes (Hybond-N, Amersham) and probed with nick-translated cDNA using [α - ^{32}P]dCTP for 2 days at 42°C as described (4). The filters were washed in $2\times$ SSC ($1\times$ SSC = 0.15 M NaCl/0.015 M sodium citrate) at room temperature for 20 min followed by two washings at 65°C for 15 min each. The washing solution was changed to $2\times$ SSC/0.1% sodium dodecyl sulfate (SDS) at 65°C for 15 min. The final wash was carried out in $0.1\times$ SSC at 65°C for 15 min. The filters were dried and exposed to Kodak X-Omat films with intensifying screens (J. V. Doehren, Elk Grove, IL) at -70°C .

RESULTS

Gene Structure and Nucleotide Sequence. The structure of the rat ICaBP gene and the flanking regions is schematically diagrammed in Fig. 1A. The gene is contained within ≈ 3 kilobases (kb) after comparison to the cDNA sequence and was found to consist of 3 exons flanked by 2 introns with the exon–intron boundary being located at the linker between the two calcium binding domains. Exon I contains only 5' untranslated sequence of the mRNA from nt 1 to 52, while the coding sequence of the mRNA is distributed between exons II and III. Exon II [144 base pairs (bp)] includes the initiation codon and coding region for the first calcium binding domain of the protein. Exon III (209 bp) bears the second calcium binding domain and the 3' nontranslated region of the mRNA including the polyadenylation signal. Fig. 1B shows the

predicted Chou–Fasman hydrophobicity plot of the ICaBP with the two calcium binding domains and the exon–intron boundary derived from the genomic sequence. The complete nucleotide sequence of the gene with 260 nt of the 5' upstream region and 1145 nt of the 3' downstream region is given in Fig. 2.

The gene shows an interesting simple sequence of 21 noninterrupted AC repeats followed by 21 noninterrupted AG repeats in the second intron at positions 697–780. Another repeat of 17 A·T pairs is also detected 981 nt downstream from the end of the third exon. Also 3' to the gene there is a stretch of alternate (RY) $_{28}$ (where R is purine and Y is pyrimidine) observed 649 nt from the gene's end. The sequence of the 5' upstream region reveals the presence of a TATA box at -29 , and two CCAAT-like boxes (CCATT) are also observed at -99 and -174 , respectively, with some A+T-rich regions. A third CCAAT-like box is also detected at -167 with the GCAAT sequence.

At the start of the first intron a 15-bp imperfect palindrome (AGGTCAGGGTGATCT) is located that shows high homology to the proposed consensus sequence for the estrogen-responsive element described by Maurer and Notides (35) and Martinez *et al.* (36) and the glucocorticoid-responsive element reported by Scheidereit *et al.* (37). The sequence stretches from nt 51 to 65 and its relationship to the estrogen and glucocorticoid responsive elements is shown in Fig. 3.

The sequences at the exon–intron boundaries are in agreement with the consensus sequences proposed by Sharp (38).

Determination of the Transcription Initiation Site of the Rat ICaBP Gene. To determine the transcription initiation site, gene primer extension analysis using oligonucleotides complementary to nt 20–39 and 387–406 of the cDNA was performed. Fig. 4 shows that a 38-nt product is detected when the first oligonucleotide was used as primer, while a 405-nt product is obtained when using the second oligonucleotide. The first nucleotide in the previously published cDNA could not be detected in the gene sequence.

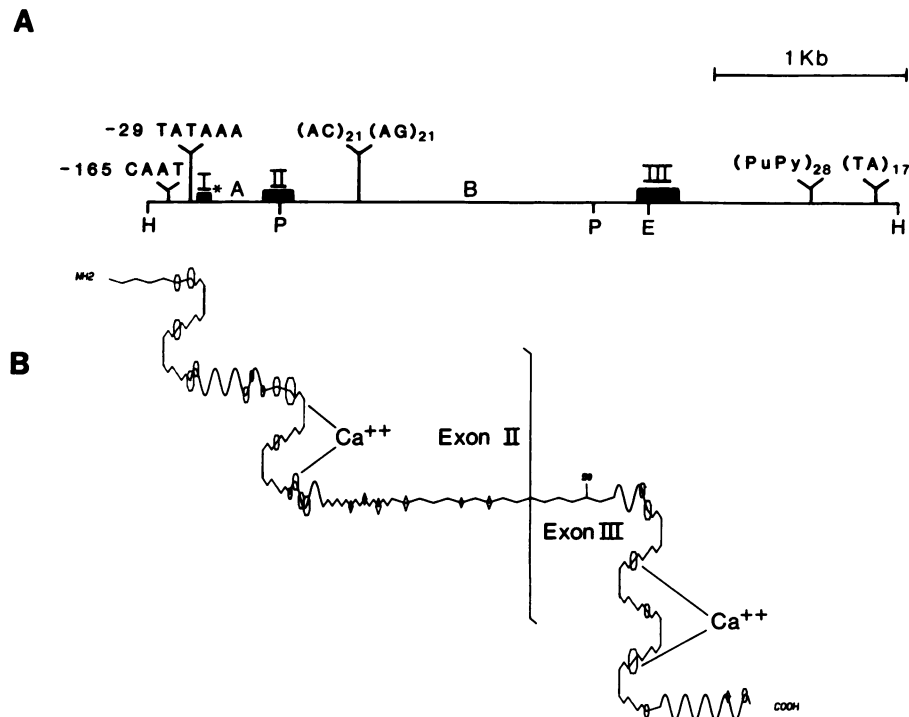


FIG. 1. (A) Schematic diagram of the rat ICaBP gene structure. Exons I–III are indicated as solid boxes. A and B denote the two introns. The location of unique sequences within the gene are shown with their nucleotide positions given from the transcription initiation site. TATA and CAAT box location are also indicated. *, A 15-bp imperfect palindrome homologous to other steroid hormone-responsive elements (see below). H, *Hind*III; P, *Pvu* II; E, *Eco*RI. (B) Predicted hydrophobicity plot of the rat ICaBP. Hydrophobic regions are shown as \diamond and hydrophilic ones as \circ . The calcium binding sites are indicated and the exon–intron boundaries of exons II and III are marked.

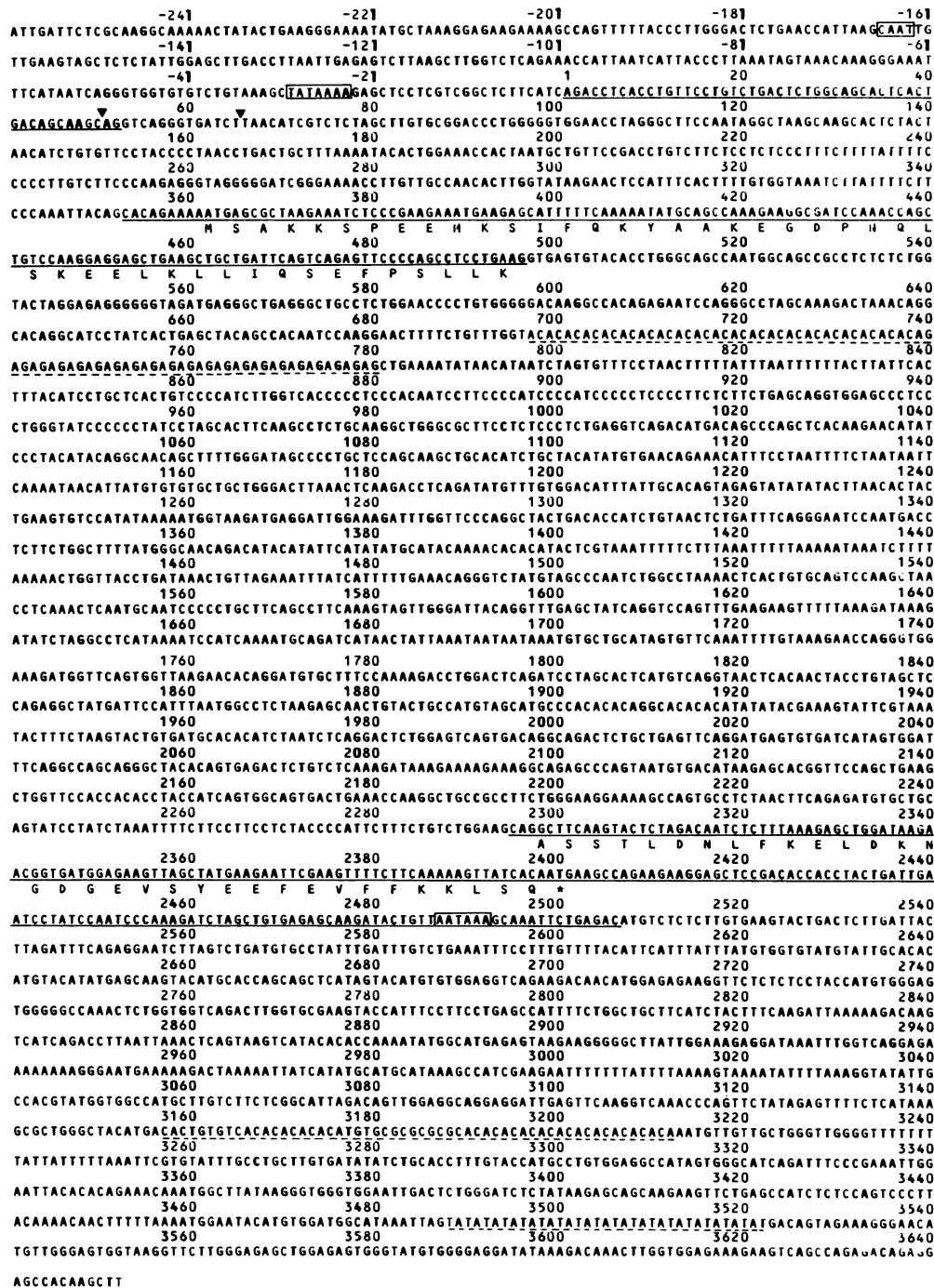


FIG. 2. Nucleotide sequence of the rat ICaBP gene. The sequence of the rat ICaBP gene including 260 nt upstream from the cap site and 1146 nucleotides of 3' downstream region is shown. TATA and CAAT box sequences and the poly(A) addition signal are indicated as boxes. Exons are underlined. The coding region marking the amino acid sequence (single-letter code) is shown under the nucleotide sequence. Unique repetitive sequences are underlined with broken lines. The 15-bp imperfect palindrome is marked by arrowheads. The probable cap site is marked as nt 1.

Genomic Southern Analysis. Fig. 5 shows the Southern analysis of rat genomic DNA digested with *Hind*III and probed with labeled rat cDNA for ICaBP. A single major band of ≈4.0 kb is detected, which is the same as that obtained for the previously isolated λ genomic clone (28). The results indicate that the gene is represented by a single copy in the genome in all tissues.

DISCUSSION

This report demonstrates that the complete rat ICaBP gene has been cloned, including some 5' and 3' flanking regions

that might be involved in regulation. The gene is contained within ≈3 kb and consists of 3 exons and 2 introns. Exons II and III represent the two different calcium binding domains, while exon I contains only noncoding sequence. The exon-intron boundary is located in between the two calcium binding domains. This structural organization is in agreement with other genes that show different protein domains represented in separate exons (39), including a calcium binding protein (40). In contrast, Simmen *et al.* (41) found three of four calcium binding domains in the chicken calmodulin gene interrupted by introns.

Interestingly, a human gene encoding a protein called

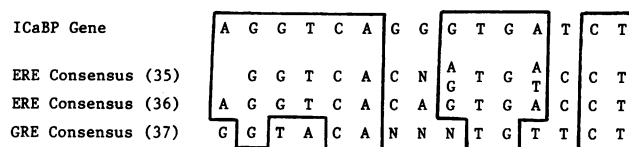


FIG. 3. Comparison of the consensus sequences described for the estrogen-responsive (ERE) and glucocorticoid-responsive (GRE) elements to the designated 15-nt sequence in the first intron of the ICaBP gene.

calcyclin reveals a very similar overall gene structure and message size (42) to the rat ICaBP gene described here. The first exon of either gene contains ≈ 50 noncoding nt, while the second exons bear ≈ 150 nt representing the first calcium

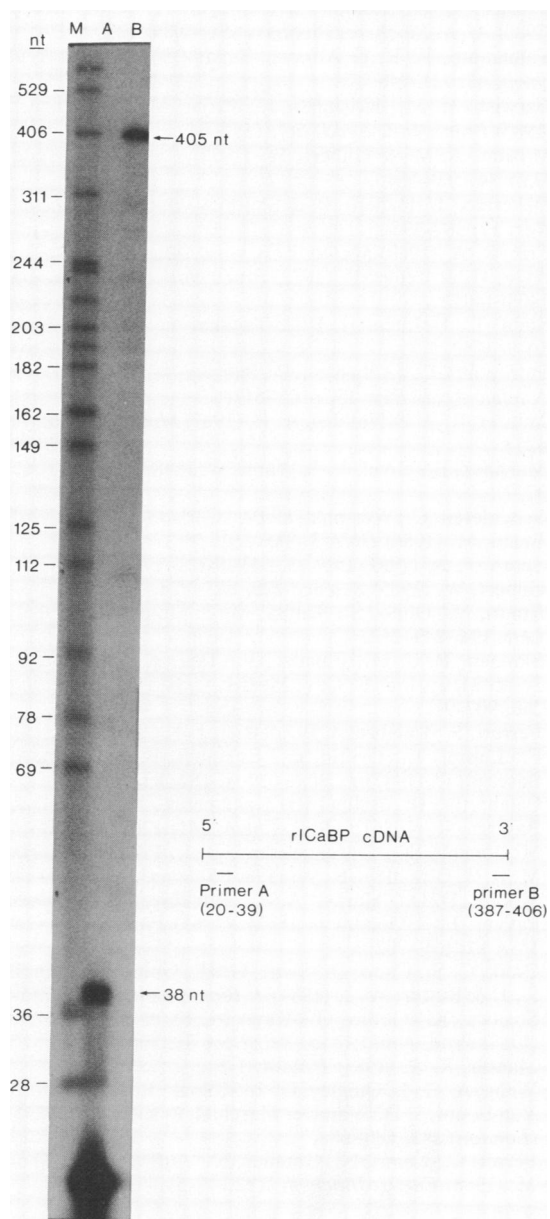


FIG. 4. Primer extension. The primer extension products of poly(A)⁺ RNA from rat intestine on a sequencing-type polyacrylamide gel are shown. The location of the two oligonucleotides used as primers A and B in the experiment are shown schematically alongside the autoradiograph. The primers were 5'-end-labeled with [γ -³²P]ATP and after hybridization extended with avian myeloblastosis virus reverse transcriptase. Plasmid pBR322 DNA digested with *Msp* I was used as molecular size marker.

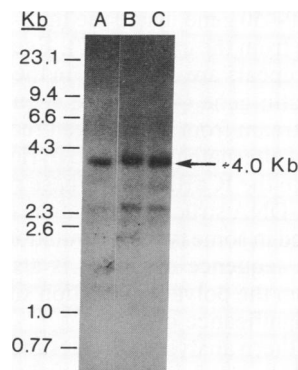


FIG. 5. Southern analysis of rat genomic DNA. DNA was prepared from intestine (lane A), kidney (lane B), and liver (lane C) of male rats. The DNA (10 μ g) was digested with *Hind*III, fractionated on 0.8% agarose, and transferred to nylon filter. The 285-nt *Eco*RI fragment of the rat ICaBP cDNA was used as probe. After hybridization, the filter was washed and subjected to autoradiography.

binding domain. In both genes, exon III includes the second calcium binding domain followed by ≈ 100 -nt 3' nontranslated region. Calcyclin is a protein that is found to be growth-regulated and inducible by platelet-derived growth factor or epidermal growth factor. The cDNA-derived protein sequence shows homology to proteins of the calcium binding protein family, in particular to the bovine S-100 and to a lesser extent to the ICaBP (43). The areas containing the two calcium binding domains reveal the highest degree of homology to the S-100 protein (43). Comparison of the nucleotide sequences of the human calcyclin and rat ICaBP gene, however, show only low homology, suggesting that if both genes are related members of the calcium binding protein gene family, their divergence has occurred in early stages of evolution.

The promoter region of the rat ICaBP gene shows a TATA box at -30, and three CAAT-like boxes at -91, -167, and -174, respectively, are observed. Two fairly large stem loops can be formed at positions -60 to -89 and -120 to -150, respectively. At the boundary of exon I and the first intron, an imperfect palindromic sequence is detected that overlaps the first exon by 2 nt and has high homology to the consensus sequences of the estrogen- and glucocorticoid-responsive elements reported by several laboratories (35-37). Within the indicated 15 nt, only two mismatches from the estrogen-responsive element consensus sequence are observed; however, there is no evidence that either estrogen or glucocorticoid exerts any regulation on the expression of this gene. The similarity of the DNA binding region of the various steroid hormone receptors including the 1,25-(OH)₂D₃ receptor has been shown (44) and the responsive elements of their target genes are quite homologous (37). It is possible that this imperfect palindrome represents the 1,25-(OH)₂D₃-responsive element that might mediate regulation of the ICaBP gene by the receptor. The location of this sequence within the first intron ≈ 60 bp downstream from the promoter region is also similar to what Moore *et al.* found in the human growth hormone gene (45). The authors report a glucocorticoid binding site about 100 bp downstream in the first intron. Regulatory sequences might be present upstream beyond the portions sequenced here or might not have been recognized by the computer. Proof of the regulatory sequences can only be established by transfection experiments.

The observation of a simple dinucleotide repeat of (AC)₂₁(AG)₂₁ in the second intron is interesting. An almost identical repeat of (AC)₂₁(AG)₂₈ has been reported in the third intron of the rat pancreatic trypsin II gene (46). Similar sequences have been found as well in the minus strand of the

rat cytochrome P-450 gene inside the first intron and 220 and 380 bp upstream from the cap site (47). Even longer stretches of dinucleotide repeats are described in the 5' flanking region of the rat prolactin gene (48). These sequences are located \approx 150 bp downstream from an estrogen receptor binding site (35). A stretch of alternate purine and pyrimidine nucleotides is also found 3' to the ICaBP gene. These repeats suggest the formation of a Z-DNA structure (49) at these positions, which might be involved in some type of regulation of the rat ICaBP. Another simple sequence of (AT)₁₇ is also detected \approx 1 kb downstream from the polyadenylation site, although we do not know the significance of it.

Genomic Southern analysis indicates the rat ICaBP gene to be represented by a single copy in the genome, since a major band at \approx 4.0 kb is detected when DNA from liver, intestine, or kidney was digested with *Hind*III and hybridized to the rat ICaBP cDNA.

The results presented in this paper constitute a major step in the understanding of the molecular steps in vitamin D transcriptional activation of its responsive genes. Due to the isolation of 7.4 and 3.2 kb of 5' and 3' flanking DNA, respectively, to the rat ICaBP gene in the original clone (28), this will allow us to investigate possible regulatory element(s) for vitamin D in this gene.

During the reviewing process of this paper, Perret *et al.* reported the isolation and sequencing of the rat ICaBP gene (50). The results obtained by that group are very similar to the comparable observations in this paper.

This work was supported by a National Institutes of Health Program Project DK-14881 and by the Harry Steenbock Research Fund of the Wisconsin Alumni Research Foundation.

- Kretsinger, R. H. (1980) *CRC Crit. Rev. Biochem.* **8**, 119–174.
- Wasserman, R. H. & Fullmer, C. S. (1983) *Annu. Rev. Physiol.* **45**, 375–390.
- Davie, M. (1981) *Biochem. J.* **197**, 55–65.
- Christakos, S., Bruns, M. E. H., Mehra, A. S., Rhoten, W. B. & Van Eldik, L. J. (1984) *Arch. Biochem. Biophys.* **231**, 38–47.
- Delorme, A. C., Danen, J. L., Acker, M. G., Ripoche, M. A. & Matheu, H. (1983) *Endocrinology* **113**, 1340–1347.
- Bruns, M. E. H., Fausto, A. & Avioli, L. V. (1978) *J. Biol. Chem.* **253**, 3186–3190.
- MacManus, J. P., Watson, D. C. & Yaguchi, M. (1986) *Biochem. J.* **235**, 585–595.
- Szebenyi, D. M. E., Oberndorf, S. K. & Moffat, K. (1981) *Nature (London)* **294**, 327–332.
- Hofmann, T., Kawakami, M., Hitchman, A. J. W., Harrison, J. E. & Dorrington, K. J. (1979) *Can. J. Biochem.* **57**, 737–748.
- Fullmer, C. S. & Wasserman, R. H. (1981) *J. Biol. Chem.* **256**, 5669–5674.
- Desplan, C., Heidmann, O., Lilie, J. W., Auffray, C. & Thomasset, M. (1983) *J. Biol. Chem.* **258**, 13502–13505.
- Leonard, W. J., Stauss, A. W., Go, M. F., Alpers, D. H. & Gordon, J. I. (1984) *Eur. J. Biochem.* **139**, 561–571.
- Bredderman, P. J. & Wasserman, R. H. (1974) *Biochemistry* **13**, 1687–1694.
- Gross, M. D., Nelsestuen, G. L. & Kumar, R. (1987) *J. Biol. Chem.* **262**, 6539–6545.
- Pansini, A. R. & Christakos, S. (1985) *Endocrinology* **117**, 1752–1660.
- Baudier, J., Glassier, N., Strib, L., Brihier, A., Thomasset, M. & Gerard, D. (1985) *J. Biol. Chem.* **260**, 10662–10670.
- Hunziker, W. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7578–7582.
- Takagi, T., Nojri, M., Konishi, K., Maruyama, K. & Nonomura, Y. (1986) *FEBS Lett.* **201**, 41–45.
- Silver, J., Russel, J. & Sherwood, L. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4270–4273.
- Rowe, D. W. & Kream, B. E. (1982) *J. Biol. Chem.* **257**, 8009–8015.
- Simpson, R. U., Hsu, T., Begley, D. A., Mitchell, B. S. & Alizadeh, B. N. (1987) *J. Biol. Chem.* **262**, 4104–4108.
- Karasawa, M., Hsoi, J., Hashiba, H., Nose, K., Tohyama, C., Abe, E., Suda, T. & Kuroki, T. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8810–8813.
- Prince, C. W. & Butler, W. T. (1987) *Collagen Rel. Res.* **7**, 305–313.
- Thomasset, M., Cuisinier-Gleizes, P., Matheu, H. & DeLuca, H. F. (1980) *Mol. Pharmacol.* **17**, 362–366.
- Thomasset, M., Parkes, C. O. & Cuisinier-Gleizes, P. (1984) *Am. J. Physiol.* **243**, E483–E488.
- Kessler, M. A., Lamm, L., Jarnagin, K. & DeLuca, H. F. (1986) *Arch. Biochem. Biophys.* **251**, 403–412.
- Yamamoto, K. R. & Alberts, B. M. (1976) *Annu. Rev. Biochem.* **45**, 721–746.
- Darwish, H. M., Krisinger, J., Strom, M. & DeLuca, H. F. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6108–6111.
- Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
- Devereux, J., Haberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
- Cox, R. A. (1968) *Methods Enzymol.* **12**, 120–129.
- Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–14012.
- Suda, T., DeLuca, H. F. & Tanaka, Y. (1970) *J. Nutr.* **100**, 1049–1052.
- Petropoulos, C. J., Yaswen, P., Panzica, M. & Fausto, N. (1985) *Nucleic Acids Res.* **13**, 8105–8118.
- Maurer, R. A. & Notides, A. C. (1987) *Mol. Cell. Biol.* **7**, 4247–4254.
- Martinez, E., Givel, F. & Wahli, W. (1987) *EMBO J.* **6**, 3719–3727.
- Scheidereit, C., Westphal, H. M., Carlson, C., Bosshard, H. & Beato, M. (1986) *DNA* **5**, 383–391.
- Sharp, P. A. (1981) *Cell* **23**, 643–646.
- Gilbert, W. (1985) *Science* **228**, 823–824.
- Emori, Y., Ohno, S., Tobita, M. & Suzuki, K. (1986) *FEBS Lett.* **194**, 249–252.
- Simmen, R. C. M., Tanaka, T., Tsui, K. F., Putkey, J. A., Scott, M. J., Lai, E. C. & Means, A. R. (1985) *J. Biol. Chem.* **260**, 907–912.
- Ferrari, S., Calabretta, B., deRiel, J. K., Battini, R., Ghezzi, F., Lauret, E., Griffin, C., Emanuel, B. S., Gurrieri, F. & Baserga, R. (1987) *J. Biol. Chem.* **262**, 8325–8332.
- Calabretta, B., Batini, R., Kaczmarek, L., deRiel, J. K. & Baserga, R. (1986) *J. Biol. Chem.* **261**, 12628–12632.
- McDonnell, D. P., Mangelsdorf, D. J., Pike, J. W., Haussler, M. R. & O'Malley, B. W. (1987) *Science* **235**, 1214–1217.
- Moore, D. D., Marks, A. R., Buckley, D. J., Kapler, G., Payvar, F. & Goodman, H. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 699–702.
- Craik, C. S., Choo, Q., Swift, G. H., Quinto, C., MacDonald, R. J. & Rutter, W. J. (1984) *J. Biol. Chem.* **259**, 14255–14264.
- Sogawa, K., Gotho, O., Kawajiri, K. & Fujii-Kuriyama, Y. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5066–5070.
- Maurer, R. A. (1985) *DNA* **4**, 1–9.
- Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., Van Boom, J. H., Van Der Marel, G. & Rich, A. (1979) *Nature (London)* **282**, 680–686.
- Perret, C., Lomri, N., Gouhier, N., Auffray, C. & Thomasset, M. (1988) *Eur. J. Biochem.* **172**, 43–51.