Sequence of Rat Intestinal Vitamin D-dependent Calcium-binding Protein Derived from a cDNA Clone

EVOlUTIONARY IMPLICATIONS*

Claude Desplan, Odile Heidmann‡, James W. Lillie§, Charles Auffray‡, and Monique Thomasset§
From the Institute National de la Santé et de la Recherche Médicale U.113, 75012 Paris and U.120, 78110 Le Vésinet, France

We have recently reported molecular cloning of the cDNA synthesized from rat duodenal mRNA-encoding intestinal calcium-binding protein (ICaBP), a vitamin D$_3$-induced protein (Desplan, C., Thomasset, M., and Moukhtar, M. S. (1983) J. Biol. Chem. 258, 2762-2765). Nucleotide sequence analysis of the longest cDNA insert (375 base pairs) permitted the assignment of 207 nucleotides of the coding region and 104 nucleotides of the entire 3'-noncoding region of the mRNA. Although the derived amino acid sequence for rat ICaBP differed from the bovine and porcine sequences by 16 and 14 residues, respectively, all the residues of each calcium-binding site followed the proposed requirements of the "EF hand" theory. In contrast, several differences found in the linker regions might explain the absence of cross-immunoreactivity between rat and porcine ICaBPs. Analysis of nucleotide sequence homologies between the coding and noncoding regions showed that the region coding for the two calcium-binding sites (I and II) was immediately followed in the noncoding region by a sequence very similar to the sequence coding for site I. This suggests that rat ICaBP mRNA contains the remains of an untranslated calcium-binding site III-like structure and that low $M_r$ ICaBP could result in early termination of the translation of a larger molecule containing four sites.

Vitamin D-dependent CaBPs† (Wasserman and Taylor, 1966; Wasserman et al., 1978) belong to the large family of intracellular proteins which bind calcium with affinity constants in the range $10^{-8}$-$10^{-9}$ m (Kretsinger, 1976). Based upon the number of their calcium-binding sites, two classes of such vitamin D-dependent proteins have been identified in mammals. A $M_r$ = 28,000 molecule found in kidney and cerebellum has four calcium-binding sites, whereas a $M_r$ = 7,500-9,000 protein mainly located in the duodenum possesses two calcium-binding domains (Wasserman et al., 1978; Baimbridge et al., 1982; Thomasset et al., 1982). On the contrary, a unique $M_r$ = 28,000 vitamin D-dependent CaBP with four calcium-binding sites has been found in high concentration in chicken intestine, kidney, and cerebellum (Wasserman et al., 1978). Chick CaBP and $M_r$ = 28,000 rat CaBP present some immunological relationships (Thomasset et al., 1982). In contrast, antibodies raised against rat intestinal CaBP do not recognize either the other mammalian intestinal CaBPs or the mammalian and avian $M_r$ = 28,000 CaBPs (Marche et al., 1977; Thomasset et al., 1982).

The synthesis of rat ICaBP in the absorptive cells of the duodenum (Marche et al., 1979) is dependent upon 1,25-dihydroxyvitamin D$_3$, the hormonal form of vitamin D$_3$ (Thomasset et al., 1979, 1980, 1982). Recently we reported the characterization of the mRNA coding for rat intestinal CaBP (Thomasset et al., 1981, 1983). We have also described the molecular cloning of a cDNA fragment, synthesized from rat duodenal mRNA-encoding rat ICaBP (Desplan et al., 1983). We now report the nucleotide sequence of the 375-bp cDNA clone and the derived amino acid sequence for rat ICaBP. We have compared the rat ICaBP amino acid sequence with porcine (Hofmann et al., 1979) and bovine (Pullmer and Wasserman, 1981) ICaBPs and integrated it into the common framework of intracellular CaBP structures, i.e. the "EF hand" theory (Moews and Kretssinger, 1975). From analysis of the nucleotide sequence homologies found within the cDNA, we also propose a general scheme of the evolution of members of the intracellular CaBP family with two or four calcium-binding sites.

MATERIALS AND METHODS

DNA-sequencing Analysis—Plasmid DNA was labeled at the 3' end either by using terminal transferase (Bethesda Research Laboratories) (Roychoudhury and Wu, 1980) and cordycepin (3000 Ci/mmol, Amersham) for the PstI and PvuII sites or by filling in the recessed ends (EcoRI, BglII, and XbaI) with the Klenow fragment of DNA polymerase I (Boehringer) (Drouin, 1980) and the appropriate 32P-labeled nucleotide followed by a chase with all four cold dNTPs. After recutting with a second restriction endonuclease, the fragments labeled at only one end were isolated on polyacrylamide or agarose gels, purified by electroelution and DEAE-cellulose chromatography. Sequencing reactions were performed according to the chemical degradation procedure of Maxam and Gilbert (1980). The products were analyzed on 20% (40 cm) and 8% (80 cm) polyacrylamide-urea thin gels according to Sanger and Coulson (1979). The nucleotide sequence was analyzed using a computer program by J. Pestell, Harvard University, Cambridge, MA.

RESULTS AND DISCUSSION

Nucletotide Sequence of cDNA-encoding Rat ICaBP—The clone containing the largest cDNA insert, termed pCaBP

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

‡ Present address, Department of Immunology, Institut Pasteur, 75015 Paris, France.
§ Present address, Department of Biochemistry and Molecular Biology, 7 Divinity Avenue, Cambridge, MA 02138.
† To whom reprint requests should be addressed at Laboratoire du Metabolisme Hydro-Mineral, Institut National de la Santé et de la Recherche Médicale U.120, 44 Chemin de Ronde, 78110 Le Vésinet, France. Recipient of grants from the Centre National de la Recherche Scientifique (Biologie Moleculaire du Gène, 034.107) and from UER Xavier Bichat, Paris VI University.

† The abbreviations used are: CaBP, vitamin D-dependent calcium-binding protein; ICaBP, intestinal CaBP; bp, base pair.
C109, was chosen for DNA sequence analysis. Fig. 1 presents the physical map obtained by restriction endonuclease analysis and the sequencing strategy. The complete sequence is presented in Fig. 2. The cloned DNA contains a 207-bp open reading frame followed by a stop codon and a 104-bp untranslated region, including the hexanucleotide AATAAA characteristic of eucaryotic mRNAs 14 bp before the poly(A) tail (Porter and Feltner, 1978). Compared with the bovine ICaBP minor component (Fullmer and Wasserman, 1981), the amino acid sequence of rat ICaBP deduced from the nucleotide sequence corresponds to amino acids 7 to 75.

**EF Hands in Amino Acid Sequence of Rat ICaBP**—Several authors (Kretzinger, 1976; Moews and Kretzinger, 1975; Goodman et al., 1979; Szebenyi et al., 1981) have postulated common features for the members of the large family of intracellular CaBPs. Each calcium-binding site is organized into a structure called an EF hand (Moews and Kretzinger, 1975), composed of a loop with two helical structures, one on either side. Several positions of such a site comprising 28 amino acids must be occupied by suitable residues (Szebenyi et al., 1981). We have compared the primary structures of rat, bovine (Fullmer and Wasserman, 1981), and porcine (Hofmann et al., 1979) ICaBPs (Fig. 3). Although the rat ICaBP sequence differs from the bovine and porcine sequences by 16 and 14 residues, respectively, all the residues present in EF hands meet the proposed requirements. The two loops are remarkably conserved, and there is no difference in the first calcium-binding site within residues 12-27, whereas only two or three differences are found in the second site within residues 53-66. In contrast, many differences are found in the linker regions and in helices at positions where any amino acid may appear. Since there is no cross-immunoreactivity between rat and porcine ICaBP (Thomasset et al., 1982), the extensive amino acid sequence homology found was rather unexpected. However, it is consistent with the structure found by Szebenyi et al. (1981), who reported that calcium-binding sites, and particularly the first site, are located inside the core of the molecule. On the contrary, regions in which there are several differences, i.e. 30-45 and 69-75, appear to be more exposed to solvent and to present more immunoreactive sites (Hopp and Woods, 1981). On the whole, the structure of the rat ICaBP reported here is in complete agreement with the EF hand theory.

**Evolution of Intracellular CaBPs**—A comparison of the rat ICaBP amino acid sequence with other published sequences of intracellular CaBPs as well as a comparison between the nucleotide sequences inside the rat cDNA clone allow us to propose a hypothesis of the molecular evolution of CaBPs. Mammalian ICaBPs (Hofmann et al., 1979; Fullmer and Wasserman, 1981), S100 brain protein (Isobe and Okuyama, 1978), parvalbumin (Berchtold et al., 1982), and skin CaBP (Rinaldi et al., 1982) contain two calcium-binding sites, whereas troponin (Collins et al., 1973), calmodulin (Dedman et al., 1982), myosin light chain (Matsuda et al., 1977), or chicken ICaBP (Bredermann and Wasserman, 1974) have four such sites. Based upon computer analysis of the amino acid sequences of these molecules, Goodman et al. (1979) proposed that the common ancestral gene for all these proteins contained four calcium-binding sites. In the ICaBPs containing four calcium-binding sites, Weeds and McLachlan (1974) noted that site I was closely related to site II and site III to site IV. In contrast, there is a weak relationship between sites I and II. This suggests that the ancestral CaBP gene arose by the duplication of a gene encoding sites I and II. Finally, mammalian ICaBPs, and S100 brain protein have a 30-residue site I instead of the 28-residue EF hand generally observed. The two additional amino acids in positions 13 or 21 do not seem to be involved in the binding of calcium (Szebenyi et al., 1981). Parvalbumin which also contains two calcium-binding sites does not present such additions (Berchtold et al., 1982).

![Fig. 1. Restriction map and DNA-sequencing strategy for the plasmid pCaBP C109. The horizontal scale indicates nucleotide bp in the same orientation as the coding sequence. The EcoRI site on the left is located in the plasmid vector. Horizontal arrows indicate the length of each restriction fragment. Both strands were sequenced one or several times by the chemical method of Maxam and Gilbert (1980) and each restriction site was overpassed in a second sequencing experiment.](image-url)

![Fig. 2. Nucleotide and deduced amino acid sequence for rat ICaBP. The nucleotide sequence was obtained as depicted in Fig. 1. The sequences are numbered beginning at the first nucleotide of codon 7, compared to bovine ICaBP (Fullmer and Wasserman, 1981). The hexanucleotide AATAAA is boxed and the restriction sites used for sequencing are indicated. The coding sequence ends with the TGA codon at position 210 (after amino acid 75).](image-url)
Sequence of Rat ICaBP from a cDNA: Evolutionary Implications

Fig. 3. Comparison of amino acid sequences of rat, bovine (Fullmer and Wasserman, 1981), and porcine (Hofmann et al., 1979) ICaBP. The sequences are presented using the single-letter amino acid code (Dayhoff et al., 1978) previously used by Szebenyi et al. (1981). Calcium-binding sites I and II are represented superposed with the parallelism of the helix-loop-helix organization. The upper line represents the EF hand test sequence (Moews and Kretsinger, 1975; Szebenyi et al., 1981) and indicates the constraints for each amino acid. *, an oxygen-containing residue (D, E, N, Q, S, T); L, a hydrophobic residue (L, V, I, K, M); G, glycine; E, glutamic acid; —, any amino acid may appear in that position. Amino acids in porcine or bovine ICaBPs are represented in order to maximize homology. Rat, bovine (Fullmer and Wasserman, 1981), and porcine (Hofmann et al., 1979) ICaBP. The sequences are presented using the single-letter amino acid code (Dayhoff et al., 1978) previously used by Szebenyi et al. (1981). Calcium-binding sites I and II are represented superposed with the parallelism of the helix-loop-helix organization. The upper line represents the EF hand test sequence (Moews and Kretsinger, 1975; Szebenyi et al., 1981) and indicates the constraints for each amino acid. *, an oxygen-containing residue (D, E, N, Q, S, T); L, a hydrophobic residue (L, V, I, K, M); G, glycine; E, glutamic acid; —, any amino acid may appear in that position. Amino acids in porcine or bovine ICaBPs are represented in order to maximize homology.

In order to understand the mechanism of evolution from ancestral forms of CaBP to rat ICaBP, we analyzed the nucleotide sequence homologies inside the coding region and also between coding and noncoding regions. The sequence encoding calcium-binding site I was compared with the region encoding site II (Fig. 4A) or the beginning of the 3'-untranslated region (Fig. 4B). The regions of the sequence corresponding to calcium-binding sites I and II are 40% homologous at the nucleotide level with 6 of 28 residues conserved (Fig. 4A). A striking homology is found when the nucleotide sequence corresponding to site I is compared to the end of the coding region (a linker not involved in calcium binding) and the beginning of the 3'-untranslated region. In order to maximize the homology of these comparisons, a limited number of insertions and deletions has to be postulated. To obtain 82% homology over a string of 41 bp (Fig. 4B), 1 base must be added at positions 216 and 232 and 3 bases deleted after the 207th base. In fact in the cDNA region encoding calcium-binding site I, this deletion corresponds to the additional amino acid necessary to validate the EF hand theory in the three known mammalian ICaBPs (residue Y at position 13 or A at position 14 (Szebenyi et al., 1981)). No homology is found between the site II sequence and the 3'-untranslated region. It appears therefore that the site II
Sequence of Rat ICaBP from a cDNA: Evolutionary Implications

**Figure 5. Model for the evolution of intracellular CaBPs.** The number of amino acids in each calcium-binding site (28 or 30 as discussed in the text) is indicated. I, II, III, and IV represent the number of calcium-binding sites.

The sequence is followed by a string of nucleotides resembling a site I sequence, suggesting that it represents part of a site III-like structure that is no longer functional since the coding frame has been interrupted at the end of the site I1 sequence. These findings suggest the following hypothesis for the evolution of CaBPs containing four calcium-binding sites.

**REFERENCES**


Moews, P. C., and Kretsinger, R. H. (1975) *J. Mol. Biol.* 91, 201–228


