BIOCHEMICAL LETTERS

Homology of cellular vitamin A-binding protein with DNA-binding proteins

Two proteins have been identified that bind vitamin A and its analogues and that facilitate the intracellular transport of these substances. These proteins are the cellular retinoic-acid-binding protein (CRABP) and the cellular retinol-binding protein (CRBP), and they are very homologous in their *N*-termini [1–3].

Vitamin A is an essential micronutrient for all vertebrate species and plays a very important role in growth, cell differentiation, and pattern formation, where it is considered to act as a morphogen [4]. The effects of vitamin A have been linked to genomic expression. Vitamin A can alter synthesis of many proteins after addition to systems *in vivo* or *in vitro* [5]. Although the molecular mechanisms involved in vitamin A actions are not understood, the interaction of vitamin A with the genome is believed to be mediated by the binding of vitamin A to proteins which facilitate its transport [5]. Therefore, elucidation of the structure and function of these proteins is of primary importance in understanding the molecular mechanisms underlying the action of vitamin A.

Data from protein sequences or DNA sequences from cDNAs corresponding to vitamin A binding proteins are now available. It has been shown that the *N*-termini of these proteins are highly homologous to myelin P2 [1]. In this communication we describe that a certain region of the vitamin A binding protein, which is located in the *N*-terminus, is highly homologous to the DNA-binding domains of several nuclear proteins. The Microgenie

Program (Beckman) was used to screen the protein data bank (GenBank) for homology with the rat cellular retinol-binding protein. As a general rule of Microgenie, the algorithm will locate regions of the data bank longer than 40 residues that are 40% homologous to protein regions of a given sequence; shorter regions having better homology will also be found. For protein alignment by similarity Microgenie uses an algorithm described by Dayhoff and others in [6]. The homologous domains are from the products of *jun* and *fos* oncogenes, the GCN4 regulatory protein of yeast, which regulates the expression of amino acid synthesizing enzymes [7], the *dnaA* protein of *Escherichia coli*, which functions in the initiation of replication [8], and the internal virion protein D of bacteriophage T7 [9], and the DNA binding sequences of λ -repressor, λ -*cro* and homoeo boxes [10,11]. Secondary structure predictions for these protein regions indicate that they are highly helix-permissive, and the same is valid for the corresponding homologous region of the retinol binding protein. Helix–turn–helix structures are believed to be the conformation that mediates the DNA-binding properties of a protein domain [7,11]. From the sequences in Fig. 1 it is calculated that for CRBP there is a 34% (46.8%) identity (similarity) with *dnaA*, 32.2% (51.6%) with virion protein D, 57.1% with *fos*, 34.7% (56.5%) with *jun*, and 27% (37.8%) with GCN4.

The above homologies suggest that vitamin A may exert its effects on genomic expression through the properties of the protein that allow it to bind to DNA. Rapid induction of the mRNA of the rat CRBP has been shown after addition of retinol [3]. This could indicate that after the addition of vitamin A or its analogue there is an induction of the proteins that bind to the vitamin

Fig. 1. Alignment of sequences from different proteins that are homologous to vitamin A-binding protein

CRBP, cellular retinol-binding protein; *dnaA*, *E. coli dnaA* protein; D, bacteriophage T7 internal virion protein D; rgb, GCN4 yeast regulatory protein; *jun*, *jun* oncogene; *fos*, *fos* oncogene; HB, homoeo box; λ -rep., λ -repressor; H2 and H3 are helix structures. Arrow heads indicate amino acids of λ -repressor and λ -*cro* which are predicted to make contact with the DNA backbone, and asterisks indicate amino acids which make major groove contacts [10]. The standard single-letter code is used for the amino acids. Identical or structurally conserved amino acids are boxed. Numbers represent the position of the first amino acid in the protein.

followed by regulation of genomic expression by specific binding of these proteins to DNA. Such a property of the vitamin A-binding proteins could have implications in the direct action of vitamin A in the fields of cellular differentiation, teratogenesis, morphogenesis and pattern formation. For example, retinoic acid has been considered to be a morphogen and its action has been shown to mimic the effects of the zone of polarizing activity (ZPA) region of the chick limb buds [4,12]. Briefly, ZPA is a region located in the posterior part of the limb buds; when it is transplanted to other sites of the limb bud, the result is the formation of duplicate limbs [13]. Retinoic acid application in the anterior part of the limb bud has been shown to have exactly the same effects as the transplantation of the ZPA region to the anterior part [4,12]. It is reasonable then to assume that the vitamin A-binding protein plays a significant role in the generation of patterns, even though nothing is known yet about the molecular mechanisms in the vitamin A-mediated generation of patterns. Some interesting clues, however, come from work on genes that control pattern formation in *Drosophila*. These genes, homoetic genes, are involved in the development of structures such as limbs, antennae, wings or the presence of cells specific to a particular lineage [14–16]. The product of the homoeotic genes are located in the nucleus and contain conserved sequences, such as the homoeo box, that interact with DNA and regulate the synthesis of other genes of the body plan as well [15]. The homologies reported in the present communication suggest that vitamin A-binding protein could be involved in pathways which lead to the regulation of genes that control pattern formation in vertebrates, possibly by binding to specific regions of DNA.

Panagiotis A. TSONIS and Paul F. GOETINCK

La Jolla Cancer Research Foundation, Cancer Research Center, 10901 North Torrey Pines Road, La Jolla, CA 92037, U.S.A.

1. Sundelin, J., Das, S. R., Eriksson, U., Rask, L. & Paterson, P. A. (1985) *J. Biol. Chem.* **260**, 6494–6499
2. Colantuoni, V., Cortese, R., Nilsson, M., Lundvall, J., Bavik, C.-O., Eriksson, U., Peterson, P. A. & Sundelin, J. (1985) *Biochem. Biophys. Res. Commun.* **130**, 431–439
3. Sherman, D. R., Lloyd, R. S. & Chytil, F. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3209–3213
4. Thaller, C. & Eichele, G. (1987) *Nature (London)* **327**, 625–628
5. Chytil, F. (1984) *Pharmacol. Rev.* **36**, 93S–100S
6. Dayhoff, M. D. (ed.) (1978) *Atlas of Protein Sequences and Structure*, pp. 345–352, National Biomedical Research Foundation, Washington DC
7. Vogt, P. K., Bos, T. J. & Doolittle, R. F. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3316–3319
8. Hansen, F. G. & Rasmussen, K. V. (1977) *Mol. Gen. Genet.* **155**, 219–225
9. Dunn, J. J. & Thompson, K., submitted to GenBank
10. Laughon, A. & Scott, M. P. (1984) *Nature (London)* **310**, 25–31
11. Pabo, C. O. & Sauer, R. T. (1984) *Annu. Rev. Biochem.* **53**, 293–321
12. Tickle, C., Alberts, B. M., Lee, J. & Wolpert, L. (1982) *Nature (London)* **296**, 564–565
13. Saunders, J. W. & Gasseling, M. T. (1968) in *Epithelial Mesenchymal Interactions* (Fleischmayer, R. & Billingham, R. E., eds.), pp. 78–97, Williams and Wilkins, Baltimore
14. Gehring, W. J. (1987) *Science* **236**, 1245–1252
15. Robertson, M. (1987) *Nature (London)* **327**, 556–557
16. Tsonis, P. A. (1987) *Trends Biochem. Sci.* **12**, 249

Received 15 September 1987