Vol. 134, No. 1, 1986 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
January 14, 1986 Pages 51-56

NUCLEOTIDE SEQUENCE OF CHICK 14K β-GALACTOSIDE-BINDING LECTIN mRNA

Yuji Ohyama*, Jun Hirabayashi*, Yuko Oda*, Shigeo Ohno**, Hiroshi Kawasaki**, Koichi Suzuki** & Ken-ichi Kasai*

* Department of Biological Chemistry, Faculty of Pharmaceutical Sciences, Teikyo University, Sagamiko, Kanagawa 199-01, Japan

** Department of Molecular Biology, The Tokyo Metropolitan Institute of Medical Science, 3-18 Honkomagome, Bunkyo-ku, Tokyo 113, Japan

Received October 31, 1985

SUMMARY: cDNA for chick 14K β-galactoside-binding lectin mRNA was cloned and the nucleotide sequence determined. The deduced amino acid sequence and the results of *in vitro* translation of its mRNA suggest that this lectin does not include any cleavable signal sequence while it exists in extracellular matrix. Comparison of the primary structures has shown that chick 14K lectin includes some regions homologous to those in discoidin I, which is also known to be located in extracellular matrix and lack signal peptide. The results imply some relation between these two lectins in spite of their great phylogenetic separation.

A family of endogenous β-galactoside-binding lectins of vertebrate tissues, having relatively low monomer molecular weights of 13,000 to 16,000 daltons, is believed to play important roles in the cellular interactions required for development and differentiation (1,2). However, their exact physiological function is not yet known because basic research on animal lectin is less advanced than that on plant lectins. We have now cloned and sequenced a cDNA for one such lectin, i.e., 14K type of chick embryo (monomer molecular weight, 14,000). The deduced amino acid sequence showed that this lectin includes some regions homologous to those in discoidin I, a developmentary regulated lectin of cellular slime mold, *Dictyostelium discoideum*. These and other structural resemblances between these two lectins should be important for an understanding of the function of β-galactoside-binding lectins of vertabrate.

MATERIALS AND METHODS

mRNA preparation and *in vitro* translation Total RNA was prepared from tarsometatarsal skin of 18-day chick embryo by the

0006-291X/86 $1.50

guanidine/CsCl method (3) and poly(A)$^+$ RNA was isolated from it by the chromatography on oligo(dT) cellulose (4). RNA was further fractionated by 5-20 % sucrose gradient centrifugation (115,000 x g, 15 hr at 20°). Then 3 μg or 0.3 μg of RNA before or after SDG centrifugation, respectively, was added to the rabbit reticulocyte lysate (Amersham) containing [^{35}S] methionine. The conditions were those recommended by the supplier. For the removal of non-specific precipitates, the reaction mixture was treated with non-immune rabbit serum before precipitation with anti rabbit IgG antiserum (Miles). The supernatant was divided into two aliquots, one of which was treated with specific antiserum and the other with non-immune serum. The preparation of antisera and the conditions for immunoprecipitation were as described before (5). Samples were separated by SDS-polyacrylamide gel electrophoresis on 12.5 % gel (6) and autoradiographed.

Cloning methods Double stranded cDNA was synthesized by the self-priming method (4) using poly(A)$^+$ RNA after SDG fractionation as the template. dC-tailed ds cDNA was annealed with Pst I-cut, dG-tailed pBR322 (New England Nuclear) and introduced into *E. coli* MM294 as described (7). For the selection of lectin cDNA clones, synthetic oligodeoxynucleotide was labeled with [γ-^{32}P] ATP (8) and hybridized at 37° with colony DNA transferred onto nitrocellulose membrane. DNA sequencing was done according to Maxam and Gilbert (8).

RESULTS AND DISCUSSION

As a first step in the cDNA cloning of chicken 14K lectin, the *in vitro* translation product of its mRNA was investigated. Since a previous study by radioimmunoassay in our laboratory had indicated that tarsometatarsal skin at the later embryonic stage contains a large amount of this lectin (5), this tissue of 18-day embryo was employed as a source of mRNA.

Fig. 1 shows the result of SDS-PAGE of the immunoprecipitate with anti 14K lectin antiserum of the translation product in the rabbit reticulocyte lysate. Only a band of 14K was observed.

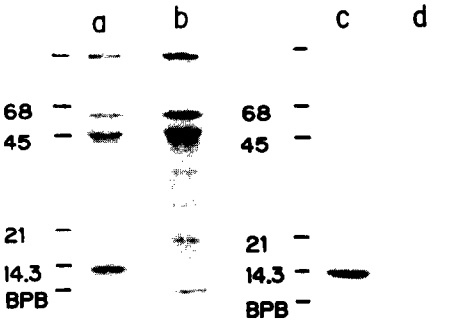

Fig. 1. SDS-PAGE of *in vitro* translation products of chick 14K β-galactoside-binding lectin mRNA. Total poly(A)$^+$ RNA (lanes a, b) or RNA fractionated by sucrose density gradient centrifugation (c, d) was translated in the rabbit reticulocyte lysate system. The products were immunoprecipitated either by anti 14K lectin antibody (a, c) or by control rabbit serum (b, d).

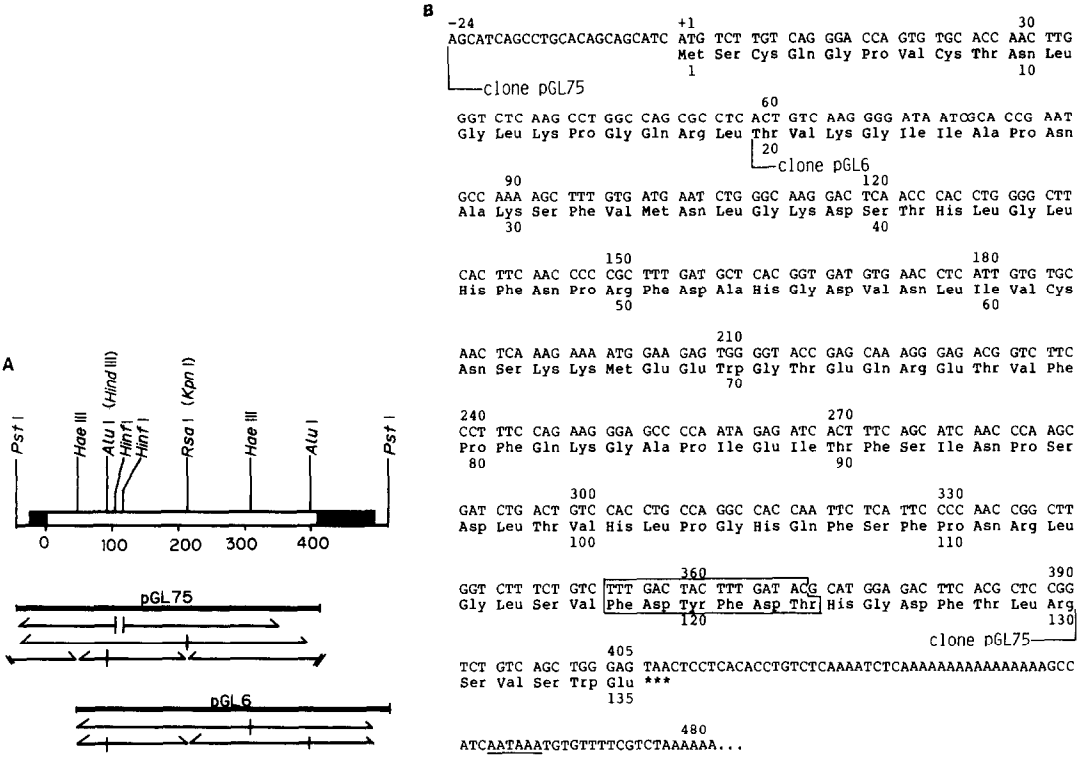

Fig. 2. a, Restriction map and sequencing strategy of chick 14K β-galactoside-binding lectin cDNA clones. Only relevant sites are indicated. The sequence of the protein-coding region is indicated by an open box, and those of 5′ and 3′ non-coding regions are by solid boxes. The direction of sequence determinations are horizotal arrows under each clone used. b, Nucleotide sequence of chick 14K β-galactoside-binding lectin cDNA. The predicted amino acid is numbered by designating the first methionine as amino acid 1. The sequence used for the synthesis of the oligodeoxynucleotide probe is blocked. The AATAAA sequence that usually precedes the poly(A) sequence is underlined.

Thus, 14K lectin appears to be produced without proteolytic processing of a precursor protein. The RNA after sucrose density gradient centrifugation was used as the template for cDNA synthesis. For the selection of recombinant clones, an oligodeoxyribonucleotide probe, $GT^{A}_{G}TC^{A}_{G}AA^{A}_{G}TA^{A}_{G}TC^{A}_{G}AA$, (17mer, 32 oligonucleotide mixture) was synthesized; it corresponds to complementary sequences for Phe-Asp-Tyr-Phe-Asp-Thr found in one of the proteolytic fragments of 14K lectin. One of the colony hybridization-positive clones, pGL75, was sequenced (the sequencing strategy is shown in Fig. 2a) and all the sequences encoding the partially determimned amino acid sequences (about 90 % of the protein sequence has been determined in our laboratory) were found in one open reading frame. Thus the nucleotide

sequence was identified as that for 14K β-galactoside-binding lectin. Fig. 2b shows the nucleotide sequences of pGL75 and its cross-hybridizing clone, pGL6. The overlapping sequences of these two clones were identical.

We have found that the N-terminal sequence of the chick 14K lectin is acetyl-Ser-Cys-Gln-Gly-Pro (the details of the protein structure will be published elsewhere). The amino acid sequence following the first Met residue of the predicted sequence is in complete agreement with this. Further, this Met should be the initiator of translation since no difference in molecular weight could be observed between the *in vitro* translation product and the mature protein. Thus mRNA for 14K β-galactoside-binding lectin should encode 135 amino acids downstream from this Met. The N-terminus is presumably acetylated after the removal of Met.

The histochemical analyses have shown that this lectin is located in the extracellular matrix of tissues (our unpublished results and ref.9). However, the results obtained here indicates the lack of any cleavable signal peptide sequence in the N-terminus of this lectin which is believed to be required for the release of extracellular or secreted protein out of the cell.

The cellular slime mold *D. discoideum* expresses two lectins, discoidins I and II, during its differentiation from the unicellular vegetative state into a multicellular organism (10). It is known that discoidin I, the predominant lectin of *D. discoideum*, does not include any cleavable signal sequence either in spite of its extracellular localization (11-14). The common feature appeared both in chick 14K lectin and discoidin I implies some structural relation between these two lectins.

When the amino acid sequence predicted here is compared with that of discoidin I, several regions appear to be homologous. Fig. 3 shows these regions. Among these, the sequences of the N-terminal regions closely resemble each other; four of six amino acids are identical besides the initiator Met.

The secondary structures predicted by the method of Chou and Fasman (15) suggest that both lectins are very rich in β sheet and poor in α helix: chick 14K, 41 % β sheet and 10 % α helix, discoidin I, 48 % β sheet and 12 % α helix. This suggests that there may be some relationship between chick 14K lectin and discoidin I. Springer *et al.* have reported that discoidin I contains the Gly-Arg-Gly-Asp sequence, which is common with the cell attachment site of fibronectin, and is considered to

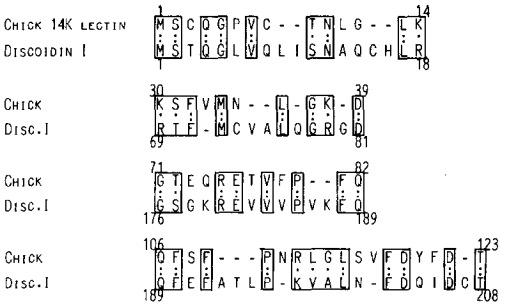

Fig. 3. Homologous sequences between chick 14K lectin and discoidin I. Upper and lower sequences show chick 14K and discoidin I, respectively. One-letter amino acid notation is used. Homologous regions are boxed; identical and homologous residues are marked by a double dot (:) and a single dot (·), respectively.

participate in the spreading of aggregation stage D. discoideum (16). A similar sequence, Gly-Lys-Asp-Ser (No.37-40), exists in chick 14K lectin. This region is located in the β-turn on the predicted secondary structure, which is also the case in discoidin I and fibronectin (17). This region of chick 14K lectin may be involved in some kind of interaction with the cell matrix of chick embryo.

Lectins that can be solubilized by lactose has been isolated in various vertebrate tissues (1,2). Those from embryonic tissues of chicken or electric organ of eel are well characterized. We have also recently purified such a lectin from human placenta (18). These lectins are thought to constitute a protein family because 1) their hemagglutinating activity is inhibited by β-galactoside-containing compounds, 2) the molecular weight of the monomer is 13 to 16 K daltons, 3) they require a sulfhydryl reagent but not any metal ion for activity. Their ubiquitous existence in vertebrates, particularly in developmental tissues, implies that they may have very important roles in the cellular interactions during development.

Chick embryo contains two isolectins, 14K and 16K type. We have extensively studied the 14K type lectin and found that the endogenous receptor for it is a proteoglycan containing a poly-lactosaminoglycan structure by using a photoaffinity labeling technique (19). This structure is implicated in the developmental processes of organs (20). The recent discovery of this structure in the carbohydrate chains of cellular fibronectin (21) has also favors the possibility that β-galactoside-binding lectins participate in cell-to-cell-interaction.

In the present study, we have elucidated the primary structure of a β-galactoside-binding lectin of a vertebrate for the first time and demonstrated a structural connection between chick 14K lectin and discoidin I, in addition to the known functional relationship. Further investigation of the structure and regulation mechanism of the lectin genes, including elucidation of the relationship with those for discoidins, seems important for an understanding of the cellular interactions occurring in the development and differentiation of higher organisms.

ACKNOWLEDGMENTS

We thank Dr Y. Emori for instruction of DNA sequencing methods and helpful discussions, Dr T. Senba of Nikkaki Ltd. for synthesis of DNA probe, Dr H. Okazaki of Seikagaku Kogyo Ltd. for permission to use the facility in some of cloning experiments, and S. Isogai and Y. Okuda for technical assistance.

REFERENCES

1. Simpson, D. L., Thorne, D. R. & Loh, H. H. (1979)Life Science 22, 727-748.
2. Barondes, S. H. (1984) Science 223, 1258-1264.
3. Chirgwin, J. M., Pzybyla, A. E., MacDonald, R. J. & Rutter, R. E. (1979) Biochemistry 18, 5294-5299.
4. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbor Laboratory, New York.
5. Oda, Y. & Kasai, K. (1983) Biochim. Biophys. Acta 761, 237-245.
6. Laemmli, U. K. (1970) Nature 227, 680-685.
7. Hanahan, D. (1983) J. Mol. Biol. 166, 557-580.
8. Maxam, A. M. & Gilbert, W. (1980) Methods. Enzymol. 65, 499-560.
9. Beyer, E. C. & Barondes, S. H. (1982) J. Cell Biol. 92, 28-33.
10. Frazier, W. A., Rosen, S. D., Reitherman, R. W. & Barondes, S. H. (1975) J. Biol. Chem. 250, 7714-7721.
11. Chang, C.-M., Reitherman, R. W., Rosen, S. D., & Barondes, S. H. (1975) Exp. Cell Res. 95, 136-142.
12. Barondes, S. H., Cooper, D. N. & Haywood-Reid, P. L. (1983) J. Cell Biol. 96, 291-296.
13. Barondes, S. H., Heywood-Reid, P. L. & Cooper, D. N. (1985) J. Cell Biol. 100, 1825-1833.
14. Poole, S., Firtel, R. A., Larmar, E. & Rowekamp, W. (1981) J. Mol. Biol. 153, 273-289.
15. Chou, P. & Fasman, G. (1978) Adv. Enzym. 47, 45-148.
16. Springer, W. R., Cooper, D. N. & Barondes, S. H. (1984) Cell 39, 557-564.
17. Pierschbacher, M. D. & Ruoslahti, E. (1984) Nature 309, 30-33.
18. Hirabayashi, J. & Kasai, K. (1984) Bichem. Biophys. Res. Commun. 122, 938-944.
19. Oda, Y. & Kasai, K. (1984) Biochem. Biophys. Res. Commun. 123, 1215-1220.
20. Muramatsu, H., Ishihara, H., Miyauchi, T., Gachelin, G., Fujisaki, T., Tejima, S. & Muramatsu, T. (1983) J. Biochem. (Tokyo) 94, 799-810.
21. Zhu, B. C.-R., Fischer, S. F., Pande, H., Calaycay, J., Shively, J. E. & Laine, R. A. (1984) J. Biol. Chem. 259, 3962-3970.