Expression of Two Different Endogenous Galactoside-binding Lectins Sharing Sequence Homology

Avraham Raz, Pnina Carmi, and Galit Pazerini
Department of Cell Biology, The Weizmann Institute of Science, P.O. Box 26, Rehovot, Israel

ABSTRACT

The endogenous $\beta$-galactoside-binding lectins of UV-2237-IP3 fibrosarcoma cells consist of two polypeptides with molecular weights of 14,500 (L-14.5) and 34,000 (L-34). Recently we constructed complementary DNA (cDNA) libraries from these cells in the expression vector Agt11 and isolated and characterized cDNA clones encoding the two lectin species (A. Raz et al., Exp. Cell Res., 173: 109, 1987). We report here the similarity in structure and gene number in genomic DNA of the fibrosarcoma cells and syngeneic normal mouse lung DNA. The expression of mRNAs hybridizing to these cDNAs was evaluated in three pairs of normal versus transformed cell variants. Messenger RNA corresponding to the M, 34,000 lectin was present more abundantly in the transformed cells than in their normal counterparts, while no difference was detected between the cell pairs with respect to mRNA levels corresponding to the M, 14,500 lectin. We have established the nucleotide sequences of the two cDNA clones. The deduced amino acid sequence of L-14.5 lectin cDNA clone (L-3) is significantly homologous to the sequence of six other L-14 vertebrate galactoside binding lectins. The sequence of the M, 34,000 lectin cDNA clone (M5) revealed that it shares a stretch of 39 amino acids with all of the L-14 galactoside binding lectins, irrespective of origin of the species, suggesting both that they are all derived from a common ancestral gene, and that this domain is necessary for similar sugar binding properties.

INTRODUCTION

Endogenous carbohydrate-binding proteins have been purified from a wide variety of normal and malignant tissues and cells. Although lectins with different sugar-binding specificities have been described, the most prevalent are those that bind $\beta$-galactosides (gal-lectins). Many of these lectins have an apparent molecular weight subunit within the range of 13,000 to 35,000. These lectins share some basic properties: their extraction into aqueous solution is facilitated in in vivo and in vitro growth, morphogenesis, differentiation, intercellular interaction, adhesion, and metastasis (1-5). Furthermore, various gal-lectins from distinct organs and species share cross-reacting immunological determinants with other gal-lectins (6-14).

Carbohydrate recognition mediated by lectins has been implicated in in vivo and in vitro growth, morphogenesis, differentiation, intercellular interaction, adhesion, and metastasis (1-5, 15, 16). Details of their molecular structure are now emerging through cloning and sequencing of several $\beta$-gal-lectins (17-21).

We have previously reported that neoplastic cells express two distinct gal-lectin polypeptide species with molecular weights of 14,500 (L-14.5) and 34,000 (L-34) (22). In order to study the structural basis for the differences between the two polypeptides and to determine whether they are structurally related to other galactoside-binding proteins, we have isolated the cDNA clone coding for these proteins from Agt11 cDNA libraries (23). Using these cDNA clones, we were able to establish that the two lectin proteins are encoded by two different genes which are not expressed in normal mouse liver (23). The present study reports on the characterization of the structural organization of the two genes; their expression was studied in transformed and normal cells. Comparison of their sequence homology with six other vertebrate L-14 gal-lectins provides the first evidence that vertebrate lectins of different molecular weights may be derived from a common ancestral gene.

MATERIALS AND METHODS

Cells and Culture Conditions. The B16-F1 melanoma cell line was obtained from Dr. L. J. Fidler (The University of Texas, M. D. Anderson Hospital and Tumor Institute at Houston, TX). Primary MDF were prepared from 12-day-old C57BL/6 mouse embryos. The UV-2237-IP3 cell line was obtained as previously described (24). 3T3 fibroblasts and SV40 polyoma virus-transformed 3T3 cells (3T3-SV40) were obtained from Dr. A. Ben-Ze'eve (The Weizmann Institute of Science, Rehovot, Israel). The NRK cells and NRK-RSV cells were obtained from Dr. B. Geiger (The Weizmann Institute of Science).

All cell lines were grown in monolayers on plastic in Dulbecco's modified Eagle's medium (GIBCO) containing 10% heat-inactivated fetal bovine serum (Bio-Lab, Israel), nonessential amino acids, and antibiotics. The cells were maintained at 37°C in a humidified atmosphere of 7% CO2 and 93% air.

RNA and DNA Blot Analysis. Ten mg of cytoplasmic RNA were fractionated by electrophoresis on 1% formaldehyde agarose gels (25) and blotted onto nitrocellulose. The filters were probed with [32P]dATP- and [32P]dCTP-labeled nick-translated pL3 (encoded for M, 14,500 lectin) and pM5 (encoded for the M, 34,000 lectin) (23) (specific activity, 2 to $5 \times 10^7$ cpm/ug; 3 x $10^7$ cpm/filter). The filters were washed twice in 2 x SSC/0.2% SDS for 30 min at room temperature and twice in 0.1 x SSC/0.1% SDS for 30 min at 50°C. High-molecular-weight DNA (20 ug) was digested with EcoRl, SacI, or PstI (Pharmacia), electrophoresed in 0.7% agarose gel, and transferred to nitrocellulose filters. The filters were hybridized with [32P]labeled probes (2 to $5 \times 10^7$ cpm/ug; 5 x $10^6$ cpm/filter) in 50% deionized formamide/5 x SSC/20 mM sodium phosphate buffer (pH 6.4)/10% dextran sulfate/100 $\mu$g/ml of denatured salmon sperm DNA for 24 h at 42°C. The filters were washed twice in 2 x SSC/0.1% SDS for 30 min at room temperature and twice in 0.1 x SSC/0.1% SDS at 60°C for 30 min.

DNA Sequencing. The EcoRl insert of pL3 was released and subcloned into the EcoRI site of the M13 mp10 and mp11 vectors. The EcoRl-Xbal, the Xbal-PstI, and PstI-EcoRI fragments of pM5 were released and subcloned into the EcoRI-Xbal, the Xbal-PstI and PstI-EcoRI sites of M13 mp10 and mp11 vectors, respectively, using the M13 cloning kit (Amersham). Dideoxy sequencing was performed using the M13 sequencing kit according to the manufacturer's instructions (Amersham).

RESULTS

DNA Blot Hybridization Analysis. We have recently shown, by RNA and DNA blot analyses of the UV-2237-IP3 fibrosar-
comica cells using two different gal-lectin cDNA clones, that the M, 14,500 and 34,000 polypeptides are encoded by two different genes (23). Since the study was only carried out on a selected fibrosarcoma cell line, UV-2237-IP3 (24), the question arose as to the genes' organization in the host of origin. To that end we performed a Southern blot analysis of genomic DNA prepared from the fibrosarcoma cells and from lungs of 6-wk-old female syngeneic C3H/HeJ mice. Hybridization of the restricted genomic DNA fragments with nick-translated pL3 and pM5 probes revealed no differences in the restriction pattern or in the genes' copy number between the tumor cells' DNA and the mouse lung DNA with either of the two probes (Fig. 1).

Nucleotide Sequence Analysis. The complete and partial nucleotide and amino acid sequences of several M, ~14,000 gal-lectins have been published recently, namely, chicken skin (17, 19), human lung (13), human placenta (20, 21), human hepatoma (18), and electric eel (21). We have previously established the identities of the two cDNA clones coding for M, 14,500 and 34,000 endogenous galactoside-specific lectins using the hybrid-selected and immunoprecipitation techniques (23). We now describe their nucleotide sequence determined by the di-deoxy termination method (26) and the identity of their relationship with the available vertebrate gal-lectin sequences.

The nucleotide sequence of the UV-2237-IP3 M, 14,500 L-3 insert is shown in Fig. 2. The sequence reveals a single open reading frame of 321 nucleotides, commencing with the ATG codon at position 15(+1) (Fig. 2). To determine whether this sequence is similar to any previously published sequences, the L-3 sequence was initially compared with the EMIB and NBRF protein and gene banks using the programs SEARCH and ALIGN, with no obvious matches being detected. The predicted amino acid sequence of L-3 was aligned with the known sequences of other M, ~14,000 gal-lectins (Fig. 3) and revealed long domains having striking homology between the mouse M, 14,500 lectin cDNA clone and the other six lectin species, with the highest homology of 69%, within the sequenced portion of the human lung lectin. The lowest level of homology (31%) was found with the eel lectin (Fig. 3). Homology between the mouse L-3 and the chicken, human placenta, and human hepatoma clones 1 and 2 was 47%, 39%, 39%, and 41%, respectively. The interhomology sequence among the non-mouse gal-lectins is within a similar range as above (Fig. 3). All 7 lectins share 20 identical residues; 28 singular residues are substituted randomly among them, and in another 20 positions they have identical substitutions which are shared by two or more species.

Fig. 2. Nucleotide sequences and predicted amino acids of L-3 cDNA clone. Nucleotides are numbered by designating the A of the first methionine as nucleotide 1.

Fig. 3. Alignment of M, 14,000 gal-lectin sequences. The sequences were obtained from mouse L-3 (Fig. 2), chicken (18), eel (22), human lung, hepatomas 1 and 2 (19), and human placentas (22) and from residue 127 till end (20). Blank spaces, nonsequenced residues; —, residues found absent during sequence alignment. Boxes enclose homologous residues. Identical substitutions shared by two or more lectins are marked with an asterisk.

The nucleotide sequence and the translated amino acid sequence of the UV-2237-IP3 M, 34,000 M5 insert are shown in Fig. 4. This cDNA insert contains an incomplete reading frame of 417 bases, followed by a stretch of 205 noncoding residues at the 3' end. The sequence contains an intact AAT AAA
polyadenylation signal but no polyadenylate tail (Fig. 4). Based on the molecular weight of 34,000, we calculated that the M5 clone is missing an upstream stretch of about 606 bases to the ATG initiation codon. A complete search for sequence homology was performed as above, but no homology with other known sequences was found. As the mouse M, 34,000 polypeptide is a gal-lectin (5, 14), we decided to align the M5 sequence with all of the available gal-lectin sequences in order to search for a homologous region based on similar sugar-binding properties. The results of such alignments showing the only stretch of homology among the eight sequences are depicted in Fig. 5. The first detected residues of homology are Asp (position 42) and Ala (position 44) (positions after the first Met of the L-14 lectins) of the electric eel L-14 followed by a His (position 46) residue which is conserved in all of the sequence segments of the gal-lectins (Fig. 5). The homology extends downstream to the Gly residue at position 84. Over this stretch of 43 amino acid residues, the interhomology sequence among the various L-14 lectins and between them and the mouse L-34 lectins is similar, >40% (Fig. 5). In addition to the 12 amino acid residues which are conserved in all of the sequenced gal-lectins irrespective of origin of the species and protein size (Fig. 5), there are 10 additional substitutions of a single amino acid which are conserved and substituted randomly in only one or 2 of the 8 gal-lectins (Fig. 5). Altogether, the 22 amino acids represent 51% of the 43 amino acid residues.

RNA and DNA Blot Hybridization Studies. We have previously shown, using immunoprecipitation analysis, that normal embryonal fibroblasts mainly express the M, 14,500 endogenous lectin species, while transformed cells coexpress both the M, 14,500 and the 34,000 lectin species (5, 14). After establishing the identity of the cDNA lectin clones, we studied the hybridization of the cDNA probes pL3 and pM5 to three pairs of normal and transformed cell variants: (a) C57BL/6 MDF-B16-F1; (b) 3T3-T3SvPy; and (c) NRK-NRK-RSV (Fig. 6). While the three normal cell variants and their transformed counterparts express almost the same level of pL3 hybridizing mRNAs, only in the 3T3 pair do the transformed cells express a 2-fold increase in mRNA expression as compared to the normal cell counterpart (Fig. 6I); markedly different levels of mRNAs hybridizing with the pM5 probes were observed between the three normal and transformed pairs (Fig. 6II). The normal cells in the three systems showed significantly less (5 – 50-fold) pM5 hybridizing mRNAs than their respective transformed cell lines (Fig. 6II), each of the probes hybridizing mRNA species corresponding to ~750 and ~1650 bp for pL3 (Fig. 6I) and pM5 (Fig. 6II), respectively, which comigrate with the UV-2237-IP3 mRNAs from which they were cloned (23).

The differential expression in the normal and transformed M, 34,000 mRNAs (Fig. 6II) raised the question as to the number of gene copies in the normal cells and their transformed counterparts. Southern blot analyses of genomic DNA of the six cell variants show no difference in the intensity or fragment pattern between each of the normal and transformed pairs (Fig. 7). The four mouse cells
exhibited a similar restriction pattern with a lower number of gene copies in the 3T3 system as compared to the C57BL/6 system (Fig. 7). A species difference in the genes’ organization is readily detected in the rat system (NRK), compared to the mouse DNAs (Fig. 7). Thus the accentuated levels of specific mRNA found in the transformed cells in comparison to the normal cell types are most likely due to augmented transcription rather than gene amplification.

DISCUSSION

Previously we identified cDNA clones coding for the M, 14,500 and 34,000 gal-lectins by using immunological cross-reactivity of the fusion proteins from the cDNA clones, and by direct immunoprecipitation of the translation polyepitides from hybrid-selected UV-2237-IP3 mRNA. In both cases the size of the specific immunoprecipitated polypeptide products corresponded to the affinity purified lectins (23). In this paper we established the identities of the cDNA clones by sequencing and comparison of predicted and protein sequences of several L-14 gal-lectins. Taken together, the data presented in this paper indicate that we have isolated authentic mouse M, 14,500 and 34,000 β-D-galactoside-binding lectin cDNA clones.

Based on two-dimensional tryptic mapping (14) and Southern blot analysis of restricted genomic DNA (23), we previously concluded that the M, 14,500 and 34,000 lectin proteins are encoded by two different genes. The present results both corroborate this conclusion and extend further to show the similarity in arrangement and number of gene copies coding for the lectin proteins between the UV-2237-IP3 tumor cells and the normal C3H mouse lung cells.

The DNA sequence and predicted amino acid sequence data of the mouse M, 14,500 cDNA clone presented here overlap with the electric eel (21), chicken (17), and four different human L-14 gal-lectins (18–21). The sequence homology among the seven L-14 lectins ranges from 34% to 70%, with no obvious preferential homology among the different species of origin; for example, the mouse lectin has 69% homology with the human lung lectin, whereas homology between the human lung and placental lectins is only 45% over the positions compared. In general, although homology extended throughout the length of the sequences, the 5' and 3' ends of the molecules displayed greater variability compared to inner stretches.

cDNA sequence analysis of the M, 34,000 lectin cDNA clone revealed that it encompasses about 40% of the predicted protein region, and the NH₂-terminal part is missing. Nevertheless, alignment analysis with the other seven L-14 gal-lectins revealed significant homology among all of them, stretching from His at position 46 of the L-14 lectin downstream to Gly at position 84. Of this, 39 residues and 12 amino acids are identically conserved with an additional eight amino acids being substituted randomly in only one or two of the eight gal-lectins, irrespective of species of origin or protein size. Could this extensive homology stretch imply that this domain is used by gal-lectins to bind β-D-galactoside? Based on spectrofluorometric studies and the effect of lactose on fluorescence of the eel lectin, it was concluded that a single tryptophan residue is present in the lactose binding site (27). pH dependence of the fluorescence, followed by sugar binding (27), and the presence of Glu residues at positions 73 and 76 prompted the suggestion that part of the β-D-galactoside-binding site resides within the amino acid residues 70 to 76 of the L-14 gal-lectin (21) which includes the tryptophan at position 70 (Fig. 5). Basically the present results do not contradict the above notion with one exception, that in mouse L-34 the glutamic acid at position 76 is substituted with lysine. We suggest expanding this domain to stretch between the histidine residue at position 46 to glycine at position 84, which includes the 12 conserved amino acid residues, because it is highly likely that this domain is important for β-D-galactoside binding, a function common to all of the eight proteins.

We have previously demonstrated that normal embryonal fibroblasts mainly express the M, 14,500 lectin species, while oncogene-transfected cell clones, derived from the normal embryonal fibroblasts, and established tumor cells abundantly coexpress both the M, 14,500 and the 34,000 gal-lectins (5, 14, 22). We show here, in both normal and transformed cells, that while there is no significant difference between the pairs in their expression of the M, 14,500 gal-lectin mRNAs, differential expression of the M, 34,000 gal-lectin mRNAs exists whereby the mRNA level in three transformed cells far exceeds that expressed in their normal cell counterparts. We have demonstrated by Southern blot analyses of genomic DNA prepared from the normal and transformed cells that the M, 34,000 gene is not transformation linked. Expression of this mRNA, therefore, must have resulted from differential gene activity. These cDNA clones, and those which we are cloning, allow detailed analyses of the differential function of the two galactoside-binding lectins in the biology of normal and tumor cells. It will also provide the means to study the molecular mechanism of sugar binding by the proteins.

ACKNOWLEDGMENTS

We thank Dr. V. Rotter and Dr. M. Oren for their critical review of the manuscript, P. Rubinstein for assistance in its preparation, and R. Karakash for handling the cells.

REFERENCES


Expression of Two Different Endogenous Galactoside-binding Lectins Sharing Sequence Homology

Avraham Raz, Pnina Carmi and Galit Pazerini


Updated version  Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/48/3/645

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/48/3/645.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.