Identification of the Metastasis-associated, Galactoside-binding Lectin as a Chimeric Gene Product with Homology to an IgE-binding Protein

A. Raz,* G. Pazerini, and P. Carmi


ABSTRACT

We report the complete primary and secondary structures of a metastasis-associated M, 34,000 galactoside-binding lectin. The polypeptide sequence (264 amino acids) was derived from the nucleotide sequence of three overlapping complementary DNA clones isolated from λgt11 and λgt10 phage libraries of UV-induced murine fibrosarcomas. Striking features of the polypeptide sequence are two distinct regions of β-sheet and globular structures at the amino and carboxy terminals, respectively. Homology search suggests that the polypeptide is a chimeric gene product formed by fusion of the 5'-end of an M, ~14,000 galactoside-binding lectin with an internal domain of the collagen α gene. Enzymatic treatment with collagenase confirmed the presence of a collagen-like structure in the polypeptide. Unexpectedly, the entire sequence is >85% homologous to a rat low affinity IgE-binding protein.

INTRODUCTION

Endogenous gal-lectins are a group of polypeptides found both on the cell surface and intracellularly (1). Although the function of these polypeptides in normal cells has yet to be established, it has been suggested that they mediate cognitive interactions between adjacent cells by binding complementary glycoconjugates and, thus, participate in morphogenesis and homeostasis (2-4). The most abundant form of gal-lectins in normal cells and tissues has a M, of ~14,000. Several gal-lectins have been recently sequenced, and the homology among those from various vertebrate species, ranging from eel to human, was found to be between 40 and 80%, suggesting a common ancestral gene (5-12).

Neoplastic transformation with concomitant expression of an additional gal-lectin of M, 34,000 (13) has been observed upon oncogene transfection of normal rat embryonal fibroblasts, which otherwise express only the M, 14,500 lectin species. The L-34 lectin is also expressed in a wide range of tumors, spontaneous, viral, UV, and chemically induced (1). Further more, among related tumor cell variants of the K-1735 melanoma, the UV-2237 fibrosarcoma and the A31-angiosarcoma, spontaneous, viral, UV, and chemically induced (1). Further more, among related tumor cell variants of the K-1735 melanoma, the UV-2237 fibrosarcoma and the A31-angiosarcoma, expression is highest in those cells that exhibit the greatest lung-colonizing (i.e., metastatic) potential (13, 14). We have, therefore, suggested that the presence of the gal-lectin on the cell surface may be related not only to neoplastic transformation but also to progression toward metastasis (1, 13, 14). Studies aimed at unveiling the function of the tumor cell surface lectin suggested that it may be involved in anchorage-independent growth regulation and could affect tumor metastasis by promoting the formation of multicell emboli in the circulation (15, 16).

We have now cloned cDNA containing the entire coding sequence of L-34. As detailed below, analysis of the lectin structure shows that the peptide is most probably a product of a chimeric gene. The entire L-34 is homologous to the rat low affinity IgE receptor.

MATERIALS AND METHODS

Cells and Culture Conditions. The UV-2240 cell line was obtained from Dr. I. J. Fidler (The University of Texas, M. D. Anderson Hospital and Tumor Institute at Houston, TX). The UV-2237-IP3 cell line was obtained as previously described (17).

Preparation and Screening of cDNA Libraries. The initial attempts to clone the L-34 cDNA from UV-2237 fibrosarcoma cells λgt11 libraries were described previously in detail (18). The presence of repeated sequences in those cDNA clones prompted the suspicion of possible mistakes in transcription. Therefore we cloned the UV-2237-IP3 fibrosarcoma and the additional murine fibrosarcoma UV-2240 into λgt10 to establish the nature of the repetitive sequences. The libraries were screened for additional clones after amplification, using 32P-labeled nick-translated pM5 clone and synthetic oligonucleotide primer (TTTGTCTAAACGATGCTCTT) as probes for DNA sequencing. cDNA restriction fragments (Fig. 1) were subcloned into the sequencing vectors M13mp10 and M13mp11 using the M13 cloning kit (Amersham). Dideoxy sequencing was performed using the M13 sequencing kit according to the manufacturer's instructions (Amersham).

Collagenase Treatment and Immunoblotting. Cell extracts equivalent to 2 x 10^6 UV-2237-IP3 cells were incubated for 2 h at 37°C in phosphate-buffered saline (pH 7.2) that contained 0.05% Triton X-100 and 2 mM phenylmethylsulfonyl fluoride, with or without 10 μg/ml collagenase (EC.3.4.24.3; Boehringer). The extracts were then separated by a reduced 15% SDS-PAGE and transferred to nitrocellulose filters. The nitrocellulose filters were incubated with monospecific antibodies raised in a rabbit against a synthetic peptide of the 172-209 region of L-34. (Fig. 1). After washing, the filters were incubated with iodinated goat anti-rabbit IgG (500,000 cpm/ml; specific activity, 11.4 μCi/μg ICN). The filters were washed, dried, and autoradiographed.

RESULTS AND DISCUSSION

Fig. 1 depicts the sequencing strategy and structure of the cDNA encoding the mouse tumor cell L-34 and gives the complete sequence of 144 5'-flanking, 792 coding, and 205 3'-flanking nucleotides. The three subclones have identical restriction patterns in the overlapping regions and hybridize with the same size mRNA of the UV-2237-IP3 cells. The 5'-nontranslated sequence is G plus C rich (86%) and the consensus initiation sequence (GGAGNNATGG) (19) surrounds the ATG(+1) codon (Fig. 1). The open reading frame encodes 264 amino acids, and the mature protein has a predicted molecular mass of 30,255. The larger apparent size on SDS/PAGE (M,
34,000) (1) is likely secondary to the unique structure of the polypeptides (see below). The 3'-nontranslated region contains an intact polyadenylation signal (AATAAA). In vitro RNA synthesis of M1 in the SP64 system (Amersham), followed by in vitro translation and immunoblotting with polyclonal antilectin antibodies, showed the appearance of a complete lectin polypeptide migrating in SDS-PAGE with M, 34,000 (not shown).

From the amino acid sequence (Fig. 1) and the predicted secondary structure (Fig. 2), the most notable features of the lectin are the two distinct half-domains of 136 and 128 amino acids for the amino and the carboxy terminals, respectively. The amino-terminal half of the protein contains predominantly α-helical structure based on the AG values for base interaction in the secondary structure (24). The carboxy-terminal half is composed of alternating reverse turns and includes both hydrophobic and hydrophilic domains (Fig. 2A), indicative of globular polypeptides. These structural arrangements could slow the migration of L-34 in SDS gels, leading to the observed size. No glycosylation or structural arrangements could slow the migration of L-34 in SDS-PAGE with M, 34,000 (not shown).

We have previously identified the putative galactoside-binding domain on L-34 (now assigned to residues 172-209), based on the carboxy-terminal partial sequence (12), which shares a conserved sequence homology with seven other low molecular weight gal-lectin species (M, ~14,000) (5-12). It is predicted that, for the sugar-binding ability, the 172-209 domain should be (a) hydrophilic in nature and (b) present on the surface of the protein. Indeed, hydrophy (20) and surface probability (21) analyses have indicated that this domain meets both expectations (Fig. 2, A and B), thus supporting the assignment of this peptide region as a galactoside binding site. In three separate experiments designed to test directly the relationship of this peptide to metastasis, we have found that a synthetic peptide corresponding to this region inhibits lung colonization of B16-F1 melanoma cells by 80% following i.v. injection of the cells with the peptide, similar to the effect of antilectin monoclonal antibodies (16).

A predicted L-34 mRNA secondary structure for the L-34 gene was obtained by a computer algorithm program (Genetic Computer Group) that calculates the most stable RNA secondary structure based on the ΔG values for base interaction in the mRNA strand. The coding region containing 792 nucleotides is shown as a hatched bar. Partial sequences hybridizing to the oligonucleotide probe indicated the nucleotide sequence in one-letter amino acid code (left).
RNA (22). The mRNA has two major stem-loop structures, which correspond to the two domains of the protein secondary structure (Fig. 2C).

A computer-assisted search of the National Biomedical Research Foundation protein data base using the FASTP program (23) identified the proline-glycine-rich amino-terminal half of the L-34 polypeptide as being homologous with the internal domain of the amino-terminus of the collagen α1 (II) chain superfamily. As an example, Fig. 3 depicts a 33.5% homology, over 124 amino acids, with the collagen α1 (II) chain of bovine cartilage (24). When the conserved amino acid substitutions are considered, the degree of homology increases to 65.5% (Fig. 3).

The collagens are a structurally related, genetically distinct, family of proteins that constitute a major component of the extracellular matrix and connective tissues of animals (25). It has been proposed that the collagen multigene family evolved from a multiexon ancestor gene, which itself evolved by multiple duplications from an ancestral unit of 54 base pairs (26). Collagens are characterized by repeats of Gly-X-Y triplets. Similar consensus repeat triplets are found in the L-34 polypeptide as being homologous with the internal domain of the L-34 molecule, with the appearance of lower M, polypeptides ranging from 20,000 to 24,000 (Fig. 5). Similar results were obtained by using previously described (13) polyclonal antibodies raised against the intact lectin molecule. The incomplete cleavage of the collagen-like structure of the lectin by the enzyme is probably due to the substitution in the collagen repeats, which interfered with the complete enzyme-substrate recognition.

The repeating conserved sequence of nine amino acid residues at the amino-terminal is likely to be important for the stability, conformation, and/or localization of the lectin in the plasma membrane. Similar repeating sequences have been found also in the nuclear antigen of Epstein-Barr virus (28) and in the circumsporozite protein of various Plasmodium species (29, 30). In both cases the repeating domains are important for human infection, and it has been suggested that they have a role in evasion of the immune system by acting as immune decoy peptides (28–30). Whether such properties can be attributed to the repeat domain of the L-34 protein will be established following the synthesis of such peptides and analysis of their effect on tumor cell growth in vivo.

Based on the homology of the carboxy-terminal sequence of L-34 with other M, 14,000 vertebrate gal-lectins, we have previously suggested that, irrespective of protein size and species of origin, all are derived from a common ancestral gene (12). The present results demonstrate that the amino-terminal half of the L-34 polypeptide is homologous to a domain of 124 amino acids of a different protein; i.e., collagen α1 (Fig. 3), and is recognized by collagenase (Fig. 5). Taken together, it thus appears that the product of the tumor cell lectin gene is a chimeric protein, resulting from fusion of the 3'-end part of the amino-terminal domain of the collagen α1 gene with the 5'-end of the M, ~14,000 gal-lectin gene. The end result is a "fused" protein which maintains its galactoside-binding properties.

There have been several other reports that describe the appearance, in cells which have undergone a carcinogenic insult, of chimeric proteins involving the fusion of an amino-terminal
receptor gene. The trk oncogene carries the amino-terminal domain of a structural gene with a putative transmembrane first methionine. Dashes indicate gaps for optimal alignment.

The L-34 is a cell surface antigen, and structural and functional analyses of the properties of the lectin suggest a membrane orientation whereby the amino-terminal is facing the cytoplasm while the carboxy-terminus, which encompasses the putative sugar-binding region, is facing outside of the cell membrane. Since the predicted amino acid sequence (Fig. 1) and the hydrophathy plot (Fig. 2A) fail to reveal an obvious signal sequence for membrane insertion near the amino terminal, it is not clear how L-34 spans the membrane. It is possible that the β-sheet structure is inserted directly through the lipid bilayer, that the internal repeats interact with another membrane-spanning integral protein, or that the lectin is associated with a native carbohydrate ligand present in the membrane surface glycoproteins. This type of membrane orientation is not unique to L-34. A different carbohydrate-binding protein, the asialofetuin receptor, displays similar membrane orientation, exposing its carboxy terminal outside the cytoplasmic membrane (35). We have previously shown that L-34 is a different gene product from the liver asialofetuin receptor (18). The present results confirm the lack of homology between the two receptors at the level of the primary amino acid sequence. However, the homology search did lead to an unpredicted and remarkable observation: the human asialofetuin receptor has considerable sequence homology with the L-34 sequence, at both the amino and carboxy terminals (35). We have previously shown that L-34 is a different, low affinity, IgE-binding protein of rat basophilic leukemia cell LE-binding protein shares 86.1% homology with the L-34 sequence, at both the amino and carboxy terminals (Fig. 6). The 86% homology present is also maintained at the level of the nucleotides (data not shown). Thus it would seem that similar molecules on the cell surface of different cell lineages may be utilized, depending on the cell type in which they are expressed, for different biological functions. This idea is supported by the consideration that the IgE molecule is a heavily glycosylated immunoglobulin.

The L-34 represents a chimeric gene product of diverse structural and functional sequences (Fig. 7), all of which may be involved in cellular recognition processes and may be expressed on the cell surface.

Our interest in the L-34 relates to its association with highly aggressive metastatic, tumor cells. Interestingly, a putative role for IgE in tumor growth has been suggested by a study in 166 nonallergic women in whom a significant positive relationship between serum IgE levels and the rate of breast cancer recurrence was found (38). Furthermore, an allergic history (6913 adults surveyed) has been reported to be associated with increased risk of subsequent malignancy (39). Additional evidence comes from animal model studies where acceleration of the growth of the Walker sarcoma in vivo was suggested to be induced by the interaction of IgE with unidentified tumor cell surface antigen(s) (40). Whether the L-34 also serves as a tumor cell receptor for IgE, which in some way facilitates proliferation, is yet to be determined.

ACKNOWLEDGMENTS

We thank Dr. G. Heppner and Dr. J. Christman for their critical review of the manuscript, N. Squiers for editing, and J. Jenkins for typing.

REFERENCES

5. Gitt, M. A., and Barondes, S. H. Evidence that human soluble β-galactoside-


Identification of the Metastasis-associated, Galactoside-binding Lectin as a Chimeric Gene Product with Homology to an IgE-binding Protein

A. Raz, G. Pazerini and P. Carmi