Primary Structure of the Soluble Lactose Binding Lectin L-29 from Rat and Dog and Interaction of Its Non-collagenous Proline-, Glycine-, Tyrosine-rich Sequence with Bacterial and Tissue Collagenase*

(Received for publication, June 7, 1993, and in revised form July 26, 1993)

Joerg Herrmann§§, Christoph W. Turck||, Robert E. Atchison‡, Margaret E. Huflejt†, Linda Poulter***, Michael A. Gitt‡, Alma L. Burlingame§, Samuel H. Barondes¶, and Hakon Leffler$$$†††

From the §§Department of Psychiatry, Langley Porter Psychiatric Institute, the ‡Department of Pharmaceutical Chemistry, and the ¶Howard Hughes Medical Institute and Department of Medicine, University of California, San Francisco, California 94143

A lactose-binding lectin from rat lung (RL-29) and a related lectin from Madin-Darby canine kidney (MDCK) cells have been analyzed with the primary goal of identifying post-translational modifications. The sequences show that RL-29 and the dog lectin are homologues of a lectin designated here as L-29 and elsewhere as CBP-35, βBP, Mac-2, or L-34. RL-29 has a 140-amino-acid COOH-terminal carbohydrate-binding domain, a 20-amino-acid NH2-terminal domain, and an intervening domain consisting of 11 repeating elements rich in Pro, Gly, and Tyr (R-domain). The dog homologue has 14 repeating elements in its R-domain explaining its larger size. The sensitivity of the R-domain to bacterial collagenase allowed us to isolate the NH2-terminal domain and show that the NH2 terminus was blocked by acetylation and, in the accompanying paper (Huflejt, M. E., Turck, C. W., Lindstedt, R., Barondes, S. H., and Leffler, H. (1993) J. Biol. Chem. 268, 26712-26718), that the NH2-terminal domain is phosphorylated. In addition, we unexpectedly found an endogenous component, resembling 92-kDa type IV collagenase, that co-purified over, the co-purification suggests a non-enzymatic interaction between 92-kDa collagenase and L-29.

L-29 is a member of the soluble lactose-binding (S-Lac) lectins, a family of lectins with calcium-independent affinity for lactose (Leffler et al., 1989). L-29 was independently discovered in man, rat, and mouse based on a variety of properties, and given different names (e.g. Mac-2, βBP, L-34, and CBP-35, see Footnote 1). However, the published cDNA sequences (Albrandt et al., 1987; Jia and Wang, 1988; Raz et al., 1989; Cherayil et al., 1989, 1990; Robertson et al., 1990; Oda et al., 1991) show that each of these proteins is either the human, rat, or mouse homolog of the same protein, here designated L-29.

The highest levels of L-29 are found in activated macrophages (Cherayil et al., 1989) and columnar epithelial cells (Oda et al., 1993; Brassart et al., 1992; Lotz et al., 1993; Lindstedt et al., 1993). The function of L-29 is unknown, but this protein has some intriguing properties; it may have both intracellular and extracellular activities. L-29 is found mainly in the cytosol and nucleus of cells (Moutsatsos et al., 1986; Lotz et al., 1993) and is synthesized without a signal peptide (Rosenberg et al., 1993) suggesting an intracellular function. Indeed, a role of L-29 in processing of RNA has been proposed (Wang et al., 1992). However, an extracellular role is suggested by the externalization of a portion of L-29 by a non-classical pathway (Lindstedt et al., 1993), and the affinity of L-29 for extracellular glycoconjugates such as laminin (Woo et al., 1990) and immunoglobulin E (Robertson and Liu, 1991). The deduced protein sequences (Albrandt et al., 1987; Jia and Wang, 1988; Raz et al., 1989; and other references given above) reveal that L-29 has the domain organization outlined in Fig. 1. The COOH-terminal 140 amino acids constitute the carbohydrate-binding domain (designated as C in Fig. 1) with sequence similarity to other S-Lac lectins. The NH2-terminal 20 residues form a distinct domain (designated as N), and the remaining residues consist of multiple copies of a characteristic Pro/Gly/Tyr-rich repeating element (designated as R). Here we describe further studies on the primary structure of rat L-29 isolated from rat lungs and dog L-29 isolated from MDCK cells.

* This work was supported by the Cigarette and Tobacco Surtax Fund of the State of California through the Tobacco-Related Disease Research Program of the University of California (to H. L.), National Institutes of Health National Center for Research Resources Grant RR01614 and National Science Foundation Biological Instrument Program (to UCSF Mass Spectrometry Facility and A. L. B.) and National Institutes of Health Grant HL36827 and National Science Foundation Grant DGB-88-10912 (to S. H. B.). 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 1794 solely to indicate this fact.

|| Present address: Glycomed, 860 Atlantic Ave., Alameda, CA 94501.

### Footnotes

1. Present address: International Development Department, Zeneca (ICI) Pharmaceuticals, Mereisde, Alderley Park, Macclesfield, Cheshire SK10 2TG, United Kingdom.

2. To whom correspondence should be addressed: Dept. of Psychiatry LPPI, Box F-0984, University of California, San Francisco, CA 94143. Tel.: 415-476-7891; Fax: 415-476-7320.

3. The abbreviations used are: L-29, an S-Lac lectin of rat L-29,000 that has been also referred to as IgE binding protein, βBP (Albrandt et al., 1987) from rat, CBP-35 (Jia and Wang, 1988), L-34 (Raz et al., 1989), and Mac-2 (Cherayil et al., 1980) from mouse, and hMac-2 (Cherayil et al., 1992) from man; S-Lac lectins, a family of soluble lectins with affinity for lactose that share some conserved sequence elements; MDCK, Madin-Darby canine kidney; LSIMS, liquid secondary ion mass spectrometry; CID, collision induced dissociation; ESI-MS, electrospray ionization mass spectrometry; HPLC, high performance liquid chromatography.
EXPERIMENTAL PROCEDURES

General—All materials, equipment, and experimental conditions were the same as described by Oda et al. (1993) unless stated otherwise.

Purification of RL-29 from Rat Lung—RL-29 was purified from rat lung by slight modifications of the procedure described by Cerra et al. (1988). First the total mixture of rat lung S-Lac lectins was isolated from frozen rat lungs (150 g, PelFreeze, AR) by affinity chromatography on lactosyl-Sepharose, exactly as described for intestinal S-lac lectins by Leffler et al. (1989). The lectin-containing fraction was dialyzed against 10 mM NaCl, 10 mM Tris-HCl, pH 7.5, and loaded onto a column of DEAE-Sepharose Fast Flow (Pharmacia LKB Biotechnology Inc.). The unbound fractions from this column contained pure RL-29 and were used for the structural analysis.

Purification of L-29 from MDCK Cells—MDCK-II cells (kindly provided by Professor Kai Simons, European Molecular Biology Laboratory, Heidelberg, Germany) were cultured under the same conditions as used by Hufeij et al. (1993). Cells grown to confluency on twelve 75-cm² tissue culture flasks (Corning) were lysed in 2 ml of 0.1% Triton X-100 (v/v), 3 mM phenylmethylsulfonyl fluoride, 10 μg/ml leupeptin, and 10 μg/ml aprotinin. The lysate was centrifuged at 10 000 × g for 15 min at 4 °C, and the supernatant was passed over a 40-ml lactosyl-Sepharose column equilibrated with phosphate-buffered saline containing 5 mM 2-mercaptoethanol and protease inhibitors (same as above). Unbound material was washed off the column with the above equilibration buffer, then with 135 mM NaCl, and 20 mM Tris-HCl, pH 7.5. The lectin was eluted with buffer containing 150 mM lactose, 50 mM NaCl, and 10 mM Tris-HCl, pH 7.5. The fractions containing pure RL-29 were used for structural analysis without further purification.

Digestion with Trypsin, Chymotrypsin, and Clostridium histolyticum Collagenases—About 600 μg of RL-29 was denatured, S-carboxymethylated, and digested with trypsin as described (Oda et al., 1993). Peptides isolated by HPLC were further digested with α-chymotrypsin (50:1; w/w), substrate/enzyme) in 200 μl of 100 mM ammonium bicarbonate, pH 8.2, at 37 °C for the times given under “Results.”

For collagenase digestion we used collagenase preparation VII (Sigma) from C. histolyticum, which probably consists of a mixture of collagenases (Bond and van Wart, 1984). About 500 μg of native RL-29 were digested with the C. histolyticum collagenases (50:1 (w/w), substrate/enzyme) in 5 ml of 60 mM sodium phosphate, 100 mM NaCl, 10 mM CaCl₂, 50 mM Tris-HCl, pH 7.4, at 37 °C for 1.5 h. About 200 μg of native dog L-29 were digested with collagenase (20:1 (w/w), substrate/enzyme) in 1 ml of 100 mM NaCl, 5 mM CaCl₂, 20 mM Tris-HCl, pH 7.5, at 37 °C for 2 h.

Mass Spectrometry—The molecular weights of peptides were determined by liquid secondary ion mass spectrometry (LSIMS) (Falicke et al., 1986). The peptide sequence was analyzed by high energy collision induced dissociation (CID) as described (Walls et al., 1990; Medzhiradszky et al., 1992). Spectra were interpreted as described (Biemann, 1988; Medzhiradszky et al., 1992).

FIG. 1. Domain structure of L-29. The COOH-domain (C) is the carbohydrate-binding domain, the R-domain is a domain consisting of Pro-, Gly-, Tyr-rich repeating elements, and the NH2-domain (N) is the first 20 amino acids. The proteolytic sensitivities of the different domains are indicated based on the results of this paper. The numbers (given as for the rat homologue) indicate the residue on the COOH-terminal side of the first and last C. histolyticum collagenase cleavage sites and the first trypsin cleavage site.

FIG. 2. HPLC of tryptic digest of RL-29. About 8 nmol of denatured and carboxymethylated protein was digested with trypsin. Half the sample was fractionated on a 4.6 × 250 mm C-18 column eluted with increasing concentrations of acetonitrile in water containing 0.1–0.08% trifluoroacetic acid. Absorbance at 214 nm was recorded at 0.1 absorbance units full scale (ordinate).

FIG. 3. Structural analysis of RL-29. The amino acid sequence encoded by the published cDNA sequence for bP (Albrandt et al., 1987) is shown. Arrowheads indicate every 20 residues. Examples of tryptic peptides (labeled with T + number of peak in Fig. 2) are shown above the sequence. Examples of peptides from the C. histolyticum collagenase digest (labeled with C + number of peak in Fig. 4a) are shown below the sequence. Peptides identified in chymotryptic digests of other peptides are labeled with Ch (+ number of peak in Fig. 4b for those derived from C-9). The differences between RL-29 and the bP sequence are indicated: Met-1 was missing and Ala-2 was acetylated in RL-29; residue 20 was Gln in RL-29 instead of Arg. Most of the peptides were identified by CID mass spectrometry whereas others (T23, T25-27, Ch9-2, Ch9-3, and C7 and C10) were identified by ESI-MS, LSIMS, and/or Edman analysis as described in the text.
Various derivatives of peptides were generated in order to enhance amino groups (Oda et al., 1987, 1993), hexylation of carboxyl groups (Falick and Maltby, 1989) and 8-28-3 Edman analysis. Molecular weight determination of the preparation of proteins (SDS-PAGE, Sigma) were used as molecular weight markers.

Electrospray ionization mass spectrometry (ESI-MS) of large proteins (SDS-PAGE, Sigma) were used as molecular weight markers.

Digestion of RL-29 with C. histolyticum Collagenases Cleaves the Repeating Element Domain at Multiple Sites and Allows Isolation of the NH2-terminal and COOH-terminal Domains—To analyze the structure in further detail, we used a preparation of C. histolyticum collagenases suggested to digest the Rat Lung Lectin RL-29 Is Identical to Rat cBP and Thus Is Rat L-29—We first analyzed the rat lung S-lact lectin RL-29 (Cerra et al., 1985) at the peptide level because its size and immunological cross-reactivity suggested that it was rat L-29. The sequences of eight tryptic peptides, determined by high energy CID mass spectrometry, were identical to those sequences, deduced from cDNA, within the COOH-domain (Fig. 1) of rat cBP (Albrandt et al., 1987) (Fig. 3). The mass of most of the other tryptic peptides (Fig. 2) matched masses expected for putative tryptic fragments of the COOH-terminal domain of cBP. For the COOH-terminal peptide (residues 242–262) the assignment was confirmed by CID-mass spectrometry of chymotryptic peptides (Fig. 3).

Some prominent late-eluting peptides in the tryptic digest (peaks 23 and 25–27, Fig. 2) were too large to be analyzed by mass spectrometry. Fractions 26 and 27 both had a blocked NH2-terminus on Edman analysis, their amino acid composition (not shown) matched the published sequence of cBP, the NHz-terminal region including residues 2–142 (Albrandt et al., 1987), and CID mass spectrometry of chymotryptic subdigests revealed the presence of peptides spanning amino acids 120–142 (Fig. 3) indicating that these fractions represent the entire NH2-terminal region including residues 2–142. Fractions 23 and 25 were similar except that Edman analysis exposed an NH2-terminal sequence identical to that starting at residue 39 in cBP (Fig. 3). These analyses revealed no differences between fractions 26 and 27 or between 23 and 25.

The identity between RL-29 and cBP, suggested by the results described above, was confirmed by analysis of peptides derived from the NH2 and R-domains (Fig. 3) and the intact COOH-domain as described below.

RESULTS

The Rat Lung Lectin RL-29 Is Identical to Rat cBP and Thus Is Rat L-29—We first analyzed the rat lung S-lact lectin RL-29 (Cerra et al., 1985) at the peptide level because its size and immunological cross-reactivity suggested that it was rat L-29. The sequences of eight tryptic peptides, determined by high energy CID mass spectrometry, were identical to sequences, deduced from cDNA, within the COOH-domain (Fig. 1) of rat cBP (Albrandt et al., 1987) (Fig. 3). The mass of most of the other tryptic peptides (Fig. 2) matched masses expected for putative tryptic fragments of the COOH-terminal domain of cBP. For the COOH-terminal peptide (residues 242–262) the assignment was confirmed by CID-mass spectrometry of chymotryptic peptides (Fig. 3).

Some prominent late-eluting peptides in the tryptic digest (peaks 23 and 25–27, Fig. 2) were too large to be analyzed by mass spectrometry. Fractions 26 and 27 both had a blocked NH2-terminus on Edman analysis, their amino acid composition (not shown) matched the published sequence of cBP, the NHz-terminal region including residues 2–142 (Albrandt et al., 1987), and CID mass spectrometry of chymotryptic subdigests revealed the presence of peptides spanning amino acids 120–142 (Fig. 3) indicating that these fractions represent the entire NH2-terminal region including residues 2–142. Fractions 23 and 25 were similar except that Edman analysis exposed an NH2-terminal sequence identical to that starting at residue 39 in cBP (Fig. 3). These analyses revealed no differences between fractions 26 and 27 or between 23 and 25.

The identity between RL-29 and cBP, suggested by the results described above, was confirmed by analysis of peptides derived from the NH2 and R-domains (Fig. 3) and the intact COOH-domain as described below.

Digestion of RL-29 with C. histolyticum Collagenases Cleaves the Repeating Element Domain at Multiple Sites and Allows Isolation of the NH2-terminal and COOH-terminal Domains—To analyze the structure in further detail, we used a preparation of C. histolyticum collagenases suggested to digest the Rat Lung Lectin RL-29 Is Identical to Rat cBP and Thus Is Rat L-29—We first analyzed the rat lung S-lact lectin RL-29 (Cerra et al., 1985) at the peptide level because its size and immunological cross-reactivity suggested that it was rat L-29. The sequences of eight tryptic peptides, determined by high energy CID mass spectrometry, were identical to sequences, deduced from cDNA, within the COOH-domain (Fig. 1) of rat cBP (Albrandt et al., 1987) (Fig. 3). The mass of most of the other tryptic peptides (Fig. 2) matched masses expected for putative tryptic fragments of the COOH-terminal domain of cBP. For the COOH-terminal peptide (residues 242–262) the assignment was confirmed by CID-mass spectrometry of chymotryptic peptides (Fig. 3).

Some prominent late-eluting peptides in the tryptic digest (peaks 23 and 25–27, Fig. 2) were too large to be analyzed by mass spectrometry. Fractions 26 and 27 both had a blocked NH2-terminus on Edman analysis, their amino acid composition (not shown) matched the published sequence of cBP, the NHz-terminal region including residues 2–142 (Albrandt et al., 1987), and CID mass spectrometry of chymotryptic subdigests revealed the presence of peptides spanning amino acids 120–142 (Fig. 3) indicating that these fractions represent the entire NH2-terminal region including residues 2–142. Fractions 23 and 25 were similar except that Edman analysis exposed an NH2-terminal sequence identical to that starting at residue 39 in cBP (Fig. 3). These analyses revealed no differences between fractions 26 and 27 or between 23 and 25.
gave a mass of the major component of 16,057 which is close to the molecular weight of 16,054 calculated for the fragment spanning residues 120-262. Two peaks (8 and 9 in Fig. 4a) had a blocked NH₂ terminus and contained the NH₂-terminal domain. Multiple smaller earlier eluting peaks contained fragments from the R-domain (some shown in Fig. 3).

The RL-29 NH₂ Terminus Is Acetylated—Fractions 8 and 9 from the collagenase digest (Fig. 4a) were treated with chymotrypsin under mild conditions. This generated three fragments from each (Fig. 4b and Fig. 3). CID mass spectrometry of fractions 8-1 and 9-1 (hexylated derivative, Fig. 5) established the modified sequence Ac-Ala-Asp-Gly-Phe. This demonstrates that the NH₂ terminus of RL-29 is lacking Met-1 of eBP and that Ala-2 is acetylated.

Edman analysis and LSIMS showed that fraction 8-2 corresponds to residues 27-47 (not shown) and fraction 9-2 to residues 27-54 (Fig. 3).

Residue 20 Is Gln in RL-29—LSIMS of fractions 8-3 and 9-3 revealed a molecular ion and sequence ions consistent with residues 6-26 of eBP, except that the mass of residue 20 was 128 (as for Gln or Lys) rather than 156 (for Arg in eBP). To distinguish the two possibilities at residue 20, the peptide was treated with reagents that incorporate an acetyl-triethylammonium group at each free amino group. Only one of these groups was incorporated (at the NH₂ terminus) showing that the peptide did not contain Lys. Therefore, residue 20 is Gln, which matches the analogous residue in human and mouse L-29.

cDNA Sequence of Dog L-29—After we found that the dog kidney MDCK cell line expresses large amounts of a 32-kDa lectin that reacts with a monoclonal anti-L-29 (anti-Mac-2) (Lindstedt et al., 1993), we screened a cDNA library from MDCK cells using human L-29 cDNA (Oda et al., 1991) as a probe. About 1% of all the plaques were strongly positive. Four clones were converted to plasmid form and sequenced. All demonstrated identical sequence, similar to L-29 from other species. The sequences included the poly(A) sequence, but the longest clone only included sequence encoding the last 104 amino acids of the protein. In order to isolate clones with more 5' sequence, we rescreened the same plates with a primer corresponding to nucleotides 418-441 (amino acids 21-28) in the human L-29. Clones which were positive for both probes were analyzed. The furthest upstream sequence (Fig. 6) obtained in these clones started with Ser at a site corresponding to Ala-14 in eBP (Fig. 3).

Identification of the NH₂- and COOH Domains of Dog L-29—We determined the remaining NH₂-terminal sequence of dog L-29 at the peptide level after sequential digestion with C. histolyticum collagenases and chymotrypsin as described for RL-29 above (Fig. 7a, top sequence). The HPLC chromatogram of the collagenase digest (shown in the accompanying paper Huflejt et al., 1993) was similar to that from RL-29. The second largest peak with a retention time similar to peak 8 in Fig. 4a was identified as the N-acetylated peptide spanning residues 2-23 by LSIMS and Edman analysis of the intact peptide and a chymotryptic fragment (residues 6-23). The reason we found only one NH₂-terminal fragment (residues 2-23), smaller than found after collagenase digestion of RL-29 (Fig. 3) is probably that the slightly modified collagenase digestion conditions applied to dog L-29 (see "Experimental Procedures") were more efficient.

The largest peak from the collagenase digest of dog L-29, with a retention time similar to peak 10 in Fig. 4a, was identified as the COOH-terminal domain spanning residues 154-296 by Edman analysis and ESI-MS.

Comparison of L-29 Sequences from Different Species—The sequence of dog L-29 is compared to the previously known sequences of human, rat, and mouse L-29 in Fig. 7. The NH₂-terminal domains and the COOH-terminal carbohydrate-binding domains are highly conserved showing about 85% identity among the different species. However, the repeating domain is more variable (see "Discussion").

Cleavage Sites for C. histolyticum Collagenases—Since our analysis of peptides generated by C. histolyticum collagenase digestion also gives some information about the specificity of this collagenase preparation we have compiled the identified cleavage sites in Table 1.

Degradation of L-29 by a Co-purifying Tissue Gelatinase
are numbered according to the complete sequence given in Fig. 2, with the starting Met considered as residue 1.

This was the fact that EDTA and compound M6001 (a metal-complexing agent) completely converted to a 1'4-kDa product. Edman analysis of this product revealed a sequence that was observed.

We examined multiple samples of L-29 and found that the degradation of the R-domain may have been due to its content of seven almost identical copies of one translational modification, since the molecular weight of the COOH-terminal domain as an intact fragment and to show that the R-domain of L-29 is sensitive to degradation by tissue gelatinases (see below).

Therefore, it is not due to different post-translational modifications because when L-29 from different tissues or cells of the same species were analyzed side by side, no differences were found (Crittenden et al., 1984; Frigeri and Liu, 1992). Instead these differences are probably due to anomalous mobility caused by the proline-rich R-domain (Albrandt et al., 1987), as found for other proline-rich proteins (Thole et al., 1991). The intraspecies differences in molecular weight reported for L-29 from rat (29,000–31,000) and mouse (32,000–35,000) are probably not due to different post-translational modifications because when L-29 from different tissues or cells of the same species were analyzed side by side, no differences were found (Crittenden et al., 1984; Frigeri and Liu, 1992). Instead these differences are probably technical, due to different conditions and standards used for SDS-polyacrylamide gel electrophoresis (as discussed in Sparrow et al., 1987).

Another potential cause for variation of the molecular weight of L-29 is proteolytic processing, either occurring naturally in the tissue, or occurring as a purification artifact. Here about 30% of rat lung L-29 lacked the first 38 amino acids. Extra bands migrating as 2-3 kDa less than the major band are often observed when cell or tissue extracts are probed with anti-L-29 (Ho and Springer, 1982; Sparrow et al., 1987; Frigeri and Liu, 1992). In the case of rat intestine, an artificial proteolysis resulted in the recovery of only the COOH-terminal domain (about 17 kDa) of L-29 (Leffler et al., 1989), and human L-29 (Robertson et al., 1990). Instead these differences are probably due to anomalous mobility caused by the proline-rich R-domain (Albrandt et al., 1987), as found for other proline-rich proteins (Thole et al., 1991). The intraspecies differences in molecular weight reported for L-29 from rat (29,000–31,000) and mouse (32,000–35,000) are probably not due to different post-translational modifications because when L-29 from different tissues or cells of the same species were analyzed side by side, no differences were found (Crittenden et al., 1984; Frigeri and Liu, 1992). Instead these differences are probably technical, due to different conditions and standards used for SDS-polyacrylamide gel electrophoresis (as discussed in Sparrow et al., 1987).

The selective sensitivity of the repeating element domain of L-29 to C. histolyticum collagenase allowed us to isolate the NH2-terminal domain as an intact fragment and to show that L-29 is post-translationally modified by cleavage of the Met-1 followed by acetylation of Ala-2, as found in other S-Lac lectins (Clerch et al., 1988; Sakakura et al., 1990). A further modification of the NH2-terminal domain, by phosphorylation, is described in the accompanying paper (Huflejt et al., 1993). No post-translational modifications were found by mass spectrometry of fragments encompassing all of the COOH-domain and most of the R-domain of rat L-29.

The post-translational modifications described above are not expected to result in a visible shift of molecular mass of L-29 on SDS-polyacrylamide gel electrophoresis. The apparent molecular mass of L-29 over the molecular weight calculated from the sequence (e.g., 31,000 and 27,500, respectively, for rat cBP, Albrandt et al., 1987) is not due to post-translational modifications, since the molecular weight of the protein translated in vitro and the protein synthesized in vivo were the same for rat L-29 (Laing et al., 1989), mouse L-29 (Cherayil et al., 1989), and human L-29 (Robertson et al., 1990).

Instead these differences are probably due to anomalous mobility caused by the proline-rich R-domain (Albrandt et al., 1987), as found for other proline-rich proteins (Thole et al., 1991). The intraspecies differences in molecular weight reported for L-29 from rat (29,000–31,000) and mouse (32,000–35,000) are probably not due to different post-translational modifications because when L-29 from different tissues or cells of the same species were analyzed side by side, no differences were found (Crittenden et al., 1984; Frigeri and Liu, 1992). Instead these differences are probably technical, due to different conditions and standards used for SDS-polyacrylamide gel electrophoresis (as discussed in Sparrow et al., 1987).

**DISCUSSION**

As an initial part of this study, we have identified a 29-kDa S-Lac lectin from rat lung (RL-29, Cerra et al., 1985) as rat L-29 and a 32-kDa S-Lac lectin from MDCK cells (Lindstedt et al., 1993) as dog L-29, and thereby distinguished them from other S-lac lectins of similar size (Hirabayashi et al., 1992; Oda et al., 1993). The sequence of RL-29 is identical to that of rat cBP (Albrandt et al., 1987), except for 1 residue which is possibly due to an error in the cDNA sequence. Hence, RL-29 is rat L-29 as defined in Footnote 1. The 32 kDa S-Lac lectin from MDCK cells is dog L-29 based on its extensive sequence similarity to L-29 from other species (Fig. 7a). It may be identical to a 35-kDa protein detected by others in MDCK cells with an antiserum raised against a 30-kDa baby hamster kidney cell lectin (Foddy et al., 1990), which has carbohydrate binding specificity (Sato and Hughes, 1992) similar to rat L-29 (Leffler and Barondes, 1986).

The deduced amino acids are numbered according to the complete sequence given in Fig. 7a, with the starting Met considered as residue 1.
in the genome because all the repeating elements of mouse L-29 are encoded by one exon (Gritzmercher et al., 1992; Rosenberg et al., 1993).

Having discussed the primary structural features of L-29, we will now discuss some functional properties of the R-domain. It has previously been shown that the NH2-terminal domain resembles a new family of collagen-like domains of some C-type lectins (Sastry and Ezekowitz, 1987). Instead, the R-domain resembles a new family of repeating sequences found in other proteins, characterized by the initiation methionines as residue 1. The intron-exon boundaries in mouse L-29 (Gritzmercher et al., 1992) are indicated by open arrows below the sequences. a, schematic of the repeating domain of L-29s.

**Carbohydrate Binding Domain**

<table>
<thead>
<tr>
<th>Animal</th>
<th>Domain</th>
<th>Number of Residues</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dog</td>
<td>N-terminal</td>
<td>153</td>
<td>PGQPGQGAP</td>
</tr>
<tr>
<td>Rat</td>
<td>N-terminal</td>
<td>119</td>
<td>FGQPGQGAP</td>
</tr>
<tr>
<td>Mouse</td>
<td>N-terminal</td>
<td>121</td>
<td>YGQPAGPLTVPYDLPLPGGWPRMLITILGTVKPNANR</td>
</tr>
<tr>
<td>Human</td>
<td>N-terminal</td>
<td>107</td>
<td>TGFAPGLTYFDLFLPGVHPMLTITGVTYPVSNR</td>
</tr>
</tbody>
</table>

**Fig. 7.** a, alignment of known L-29 sequences. The repeating elements have been categorized into seven types. When more than one copy of one of these types occurs within a sequence, the copies are shown stacked (aligned to each other) with the top copy coming first in the sequence. Thus, dog L-29 has seven copies of repeating element type III: first PGQPGQGAP followed by six copies of PGQPGQGAP. The numbers given above the top sequence indicate positions where the sequences differ, showing the number of different residues (a gap counting as one alternative) occurring at the indicated position. The number of the first residues of the carbohydrate-binding domains are indicated assuming the initiation methionines as residue 1. The intron-exon boundaries in mouse L-29 (Gritzmercher et al., 1992) are indicated by open arrows below the sequences. b, schematic of the repeating domain of L-29s.

lagen (Raz et al., 1989) and other proline-rich proteins (Jia and Wang, 1988), the lectin lacks the repeating sequence that would allow it to form a collagen-like triple helix (Albrandt et al., 1987). In this regard it is different from the collagen-like domains of some C-type lectins (Sastry and Ezekowitz, 1987). Instead, the R-domain resembles a new family of repeating sequences found in other proteins, characterized by a conserved aromatic residue (mainly Tyr) combined with at least 20% Pro and Gly, Thole et al., 1990), and which have, in the case of synexin VII, been proposed to be involved in self-aggregation (Zhang-Keck et al., 1993). These sequences are found in cystosolic domains of synexin, synaptophysin, RNA polymerase II, octopus rhodopsin, and the plant proteins gliadin, gluten, and hordein (Matsushima et al., 1990), and a
The number of the residue (for rat L-29) on the NH2-terminal side of the cleavage site is given.

In this regard it was interesting to find that an endogenous component with gelatinase activity co-purified with the lectin in the absence of calcium and degraded the lectin in the presence of calcium. The molecular mass of this component and its sensitivity to a specific inhibitor are consistent with it being the 92-kDa type IV collagenase (Fisher et al., 1989; Librach et al., 1991) which is known to be active against non-triple helical substrates (Senior et al., 1991).

The 92-kDa type IV collagenase has a fibronectin-like gelatin-binding domain (Wilhelm et al., 1989), which is different from the active site (Collier et al., 1992). This gelatin-binding domain may be responsible for binding the R-domain of L-29, resulting in co-purification of the lectin and enzyme without degradation of L-29.

The interaction between L-29 and tissue collagenases (gelatinases) in vivo requires that they are present in the same tissue and subcellular compartment. Since both L-29 (Flotte et al., 1983; Nibbering et al., 1987) and 92-kDa collagenase (Welgus et al., 1990) are associated with alveolar macrophages, and both may be externalized (the collagenase by a classical secretory pathway and L-29 by a non-classical pathway (Cherayil et al., 1989)), they may interact in this way in vivo. L-29, added externally to cultured muscle cells, which secrete a 72-kDa gelatinase, was found to be degraded; L-29 secreted by MDCK cells was not degraded (Lindstedt et al., 1993), possibly because these cells, under the culture conditions used, did not secrete a gelatinase.

Without further research, it is only possible to speculate about the functional consequences of the interaction of L-29 with collagenases. The degradation of the R-domain may result in down-regulation of L-29's agglutinating ability; another possible effect is release of chemotactic peptides since certain R-domain sequences resemble known chemotactic peptides from elastin (Mecham et al., 1989). The non-enzymic interaction may cross-link collagenase to the macrophage surface or extracellular matrix via L-29, possibly important to limit its dispersion. In conclusion, the results add support for the hypothesis, presented in the Introduction, that L-29 has both cytosolic and extracellular functions. The NH2-terminal acetylation is characteristic for cytosolic proteins and the R-domain resembles Pro-rich domains of other cytosolic proteins (Matsushima et al., 1990). However, the R-domain was also sufficiently gelatin-like to interact with gelatinases, typically found extracellularly.

REFERENCES


FIG. 8. Gelatin zymogram of a RL-29 fraction purified from rat lungs. The sample was analyzed on a 10% polyacrylamide gel containing 1 mg/ml gelatin. After electrophoresis the gel was incubated and then stained with Coomassie Blue. Gelatinase activity, revealed as a lighter band against a dark background, was observed at 92 kDa. The arrowhead to the left indicates the mobility of 92-kDa type IV collagenase from a sample of first trimester trophoblast cell medium (Fisher et al., 1989). The arrowheads to the right show the mobilities of prestained molecular weight markers (Sigma).