THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 269, No. 50, Issue of December 16, pp. 31770–31776, 1994
Printed in U.S.A.

An Adherens Junction Protein Is a Member of the Family of Lactose-binding Lectins*

(Received for publication, August 2, 1994, and in revised form, October 20, 1994)

Marie L. Chiu‡, David A. D. Parry§, Steven R. Feldman‡¶, David G. Klapper‖, and Edward J. O'Keefe‡**

From the Departments of ‡Dermatology and ‖Microbiology and Immunology, University of North Carolina, Chapel Hill, North Carolina 27514 and §Department of Physics and Biophysics, Massey University, Palmerston North, New Zealand

We previously described a pig junction protein of M_r 37,000 found in oral epithelium but not in epidermis, limited to suprabasal cells, and colocalizing by immunofluorescence with adherens junction proteins. A 1.1-kilobase pair cDNA of the 37-kDa protein yielded an open reading frame encoding a 323-amino acid protein of 35,852 Da, and Northern analysis demonstrated a band of 1.2 kilobases in tongue RNA. Secondary structure predictions indicate that the 37% identical 16–17-kDa N- and C-terminal domains form β-sheet-rich barrels linked by a compact proline-rich segment. The protein is 72% identical in amino acid sequence and shares symmetrical two-domain structure with L-36, a lectin of unknown function from rat intestine, indicating that the 37-kDa protein is the porcine form of L-36. Of the homologous lactose binding lectins known, two others, invertebrate lectins, share this symmetrical structure. Expression of the C-terminal domain of the pig lectin in bacteria yields a lectin which binds lactosyl-Sepharose, and binding is inhibited by lactose. The expressed protein binds a glycoprotein of 120 kDa from pig tongue epithelium on Western blots, and this is also inhibited by lactose. The findings suggest that the lectin function may be involved in the assembly of adherens junctions.

Soluble lactose-binding proteins found in numerous vertebrates can be purified on lactose-derivatized affinity columns (1–4) and do not require Ca^{2+} to agglutinate cells (5). Unlike membrane proteins, which required detergents for solubilization, these proteins were found to be soluble in cells and tissues disrupted in the presence of reducing agents and lactose. On the basis of sequence data, they have recently been divided into three groups according to size and homology: an M_r 14,000–16,000 group usually found as noncovalent dimers, an M_r 29,000–35,000 group, and an M_r 32,000–36,000 protein found in rat intestine (6), of which the only forms described in the literature are in rat intestine and in *Caenorhabditis elegans*.

We have described a 37-kDa epithelial protein from pig tongue that appears to be a component of adherens junctions and is unique in its restriction to limited regions and layers of stratified epithelium. The protein was present in all regions of pig oral epithelium examined but was detectable by immunofluorescence only in suprabasal cells and appears to be excluded from most regions of epidermis. We constructed a cDNA library from pig tongue mRNA and cloned and sequenced the protein. The data indicate that this epithelial junction protein is a member of the family of lactose-binding lectins and is homologous with lectins from rat intestine and two invertebrates.

MATERIALS AND METHODS

Sequence Analysis of Tryptic Peptides—The 37-kDa protein purified according to Chiu *et al.* (7) was reduced and radioalkylated with [^{3}H]iodoacetic acid, desalted, and digested with 1% (w/w) L-1-tosylamido-2-phenylethyl chloromethyl ketone-treated trypsin (Worthington) for 4 h in 1% ammonium bicarbonate. The digestion mixture was lyophilized and resolubilized in 0.1% trifluoroacetic acid, and peptides were separated by HPLC[1] (Waters Associates, Milford, MA) using a 300-Å C_{18} 4.6 × 150-mm W-Porex column (Phenomenex, Torrance, CA). The gradient consisted of 0.1% trifluoroacetic acid with increasing acetonitrile over 60 min at 0.5 ml/min. The eluate was monitored simultaneously at 214 and 254 nm, and peptides were collected in 1.5-ml polypropylene tubes and immediately frozen and stored at −20 °C. Peptides were applied to polybrene-treated glass fiber filters and sequenced in an Applied Biosystems model 470A sequencer modified with on-line HPLC analysis (Waters Associates).

cDNA Cloning—Preparation of an oligo(dT) primed λgt11 pig tongue cDNA library, cDNA cloning, Northern analysis, and isolation of the fusion protein were performed by standard methods (8). The amplified library was screened with a polyclonal antibody (7) against the purified 37-kDa protein using *Escherichia coli* strain Y1090 as host and ^{125}I-labeled protein A (labeled with chloramine T according to Hunter and Greenwood (9)) and autoradiography for detection of clones expressing the protein. Positive plaques were purified and subcloned into Bluescript (Bluescript KS, Stratagene, La Jolla, CA), and cDNA was sequenced by the dideoxy chain termination method in both directions with Sequenase (U. S. Biochemical Corp., Cleveland, OH).

Northern Analysis—RNA was extracted from keratomed pig tongue epithelium with guanidinium thiocyanate and purified on a CsCl gradient (8), electrophoresed through agarose gels containing formaldehyde, and transferred to a nylon membrane (Zeta-Probe, Bio-Rad). The 1.1-kbp cDNA obtained with the antibody to the purified 37-kDa protein was labeled with ^{32}P by random priming. The blot was prehybridized (5 min, 43 °C) and hybridized with labeled probe (6 × 10^6 cpm, 23 h, 43 °C) according to the manufacturer's instructions (formamide protocol), washed briefly in 2 × SSC (24 °C), twice in 2 × SSC, 0.1% SDS (15 min, 24 °C), and twice in 0.05 × SSC, 0.1% SDS (15 min, 65 °C). X-ray film was exposed for 24 h.

Expression of a Subdomain of the Protein—The region of the cDNA coding for amino acids 176–323 was synthesized by PCR using primers containing an *Eco*RI site (underlined) 5′-ATATCCGAATTCCCGTGTAACCTGCCATGCAT-3′ and 5′-ATATATGAATTCGATCTGGACGTAGGACAAGGT-3′. PCR was performed with 1-min denaturation at 94 °C, 2-min annealing at 60 °C, and 2-min synthesis at 72 °C with 40 cycles. The 100-μl PCR mixture contained 0.25 μM of each primer, 15 ng of purified full-length cDNA of the 37-kDa protein ligated into the Bluescript vector, 0.2 mM of each deoxynucleoside triphosphate, *Taq* polym-

* 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 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) X79303.

¶ Current address: Dept. of Dermatology, Bowman Gray School of Medicine, Winston-Salem, NC 27157.

** To whom correspondence should be addressed: 137 NCMH, Dept. of Dermatology, University of North Carolina, Chapel Hill, NC 27514. Tel: 919-966-5057; Fax: 919-966-4242.

[1] The abbreviations used are: HPLC, high performance liquid chromatography; PCR, polymerase chain reaction; kbp, kilobase pair(s).

```
   1 gtcctgcggctcttctcacgggagaagccaccagcccagcctctcaagatggccttcgtc  60
                                                  M  A  F  V    4

  61 cctgcaccaggctaccagcccacctacaatcccacgctgccctactacaagcccatccca 120
   5 P  A  P  G  Y  Q  P  T  Y  N  P  T  L  P  Y  Y  K  P  I  P   24

 121 ggcggtctccgggtcggaatgtccgtttacatccaaggagtggccaacgagcacatgaag 180
  25 G  G  L  R  V  G  M  S  V  Y  I  Q  G  V  A  N  E  H  M  K   44

 181 aggttcttcgtgaacttcgtggtggggcagggcccgggggcggatgtcgccttccacttc 240
  45 R  F  F  V  N  F  V  V  G  Q  G  P  G  A  D  V  A  F  H  F   64

 241 aatcctcgcttcgatggctgggacaaggtggtcttcaactcgcagcaggacggcaagtgg 300
  65 N  P  R  F  D  G  W  D  K  V  V  F  N  S  Q  Q  D  G  K  W   84

 301 ggcaacgaggagaagaagaggagcatgcccttccgcaaggcgcccgccttcgagctggtc 360
  85 G  N  E  E  K  K  R  S  M  P  F  R  K  A  P  A  F  E  L  V   104

 361 atcatggtcctgcccgagcactacaaggtggtggtaaacggtgatcccttctatgagttt 420
 105 I  M  V  L  P  E  H  Y  K  V  V  V  N  G  D  P  F  Y  E  F   124

 421 gggcaccggatcccagtccagttggtcacccacctgcaagtggatggggacctgacgctt 480
 125 G  H  R  I  P  V  Q  L  V  T  H  L  Q  V  D  G  D  L  T  L   144

 481 caatcaatcaacttcatcggaggccagcccgcccccagcccgggacccatgcctaatccg 540
 145 Q  S  I  N  F  I [G  G  Q  P  A  P  S  P  G  P  M  P  N  P   164

 541 gggtacccaggtcctggaaagcacaaccaacagccgtgtaacctgccatgcatggaggga 600
 165 G  Y  P  G  P  G  K  H  N  Q  Q] P  C  N  L  P  C  M  E  G   184

 601 gccccaaccttcaacccgcctgtgccatataagacgagactgcaagggggccttgtggcc 660
 185 A  P  T  F  N  P  P  V  P  Y  K  T  R  L  Q  G  G  L  V  A   204

 661 cgaagaaccatcgtgatcaagggctatgtgcccccctcgggcaagagccttgtcatcaac 720
 205 R  R  T  I  V  I  K  G  Y  V  P  P  S  G  K  S  L  V  I  N   224

 721 ttcaaggtgggctcctcagggacgtggctttgcacatcaaccccccgcctgaccgagggc 780
 225 F  K  V  G  S  S  G  D  V  A  L  H  I  N  P  R  L  T  E  G   244

 781 atcgtggttcggaacagctatctgaatggcaagtggggagccgaggagaggaagagctcc 840
 245 I  V  V  R  N  S  Y  L  N  G  K  W  G  A  E  E  R  K  S  S   264

 841 ttcaacccgtttgctcccggacagtacttcgatctgtccattcgctgtggcttggatcgc 900
 265 F  N  P  F  A  P  G  Q  Y  F  D  L  S  I  R  C  G  L  D  R   284

 901 ttcaaggtttacgccaatggccagcacctcttcgacttctcccatcgcctctcgaacttc 960
 285 F  K  V  Y  A  N  G  Q  H  L  F  D  F  S  H  R  L  S  N  F   304

 961 caaggggtggacacactggagatccagggcgatgtcaccttgtcctacgtccagatctga 1020
 305 Q  G  V  D  T  L  E  I  Q  G  D  V  T  L  S  Y  V  Q  I  *   324

1021 tctattcctggggccataacccttgggcgcacagaggaagagcctgtcggactcccttct 1080

1081 aagcctctaataaaattaataaactgccaaaaaaaaaaaaaaaaa 1125
```

FIG. 1. **Nucleotide sequence and deduced amino acid sequence of the 37-kDa protein (pig L-36).** The sequence obtained from two clones by sequencing in both directions was translated with the "Gap" program of the GCG software package (31). *Bold type*, open reading frame; *underlined regions*, tryptic peptides sequenced from the purified protein; bracketed residues in italics (amino acids 151–175), linking domain. GenBank™/EMBL accession no. X79303.

TABLE I
Comparison of experimental and predicted amino acid compositions

Amino acid	Experimental[a]	Predicted[b]
	mol %	
Asx	10.0	9.9
Glx	11.7	8.7
Ser	6.4	5.0
Gly	13.0	10.8
His	1.3	2.8
Arg	4.9	4.6
Thr	3.2	3.1
Ala	6.0	4.3
Pro	10.5	10.2
Tyr	2.2	4.3
Val	9.0	9.6
Met	1.7	2.2
Ile	3.9	4.3
Leu	6.4	6.5
Phe	5.6	6.8
Lys	4.6	5.0
Cys	ND[c]	0.9
Trp	ND[c]	0.9

[a] From Chiu *et al.* (7).
[b] From the PeptideSort program of GCG software.
[c] ND, not done.

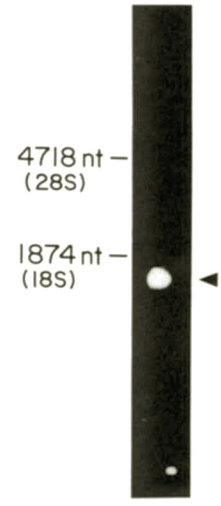

FIG. 2. **The 37-kDa protein cDNA clone identifies RNA of 1.2 kbp.** RNA was prepared from pig tongue epithelium and probed with 32P-labeled 1.1-kbp cDNA as described under "Materials and Methods." *Arrowhead*, region identified by probe. *nt*, nucleotide.

erase buffer with 2.5 mM $MgCl_2$, and 0.75 unit of *Taq* DNA polymerase (Promega, Madison, WI). The resulting DNA was digested with *Eco*RI and ligated into the pGEX-1λT expression vector (Pharmacia Biotech Inc.) as described elsewhere (8). Competent *E. coli* strain DH5α cells were transformed with the ligation product, and the expressed 15-kDa protein (designated *domain II*) fused with glutathione *S*-transferase (designated as *GII*) was isolated on glutathione agarose (8). The correct orientation of the insert was determined by reactivity of expressed protein with antibody to the 37-kDa protein on immunoblots. Glutathione *S*-transferase was prepared from competent DH5α cells transformed with pGEX-1λT and purified on glutathione agarose.

Binding of Protein to Lactosyl-Sepharose—Lactosyl-Sepharose was prepared according to Levi and Teichberg (10). To 85 μl of lactosyl-Sepharose equilibrated with MEPBS-Tw (PBS containing 2 mM EDTA, 4 mM β-mercaptoethanol, and 0.03% Tween 20), 55 μl of the same buffer containing 10 μg of glutathione *S*-transferase, GII, or glutathione *S*-transferase plus GII was added and incubated at 24 °C for 1 h on a rotator. Tubes were centrifuged and pellets washed twice with 0.8 ml of ice-cold MEPBS-Tw and once with ice-cold MEPBS. To the supernatant, 0.25 volume of 5 × concentrated Fairbanks sample buffer (11) was added, and to washed pellets, 45 μl of 2 × concentrated sample buffer was added, and the tubes were heated at 80 °C for 4 min and analyzed by SDS-polyacrylamide gel electrophoresis (11) using a 1.5-mm 3.5–17% gradient gel (12).

Inhibition of Binding of Protein to Lactosyl-Sepharose—Twenty-four μl of MEPBS containing 8.5 μg of GII was incubated at 24 °C in the absence (A) or presence of 0.2 M lactose (B) or 0.2 M sucrose (C) for 1 h on a rotator. Lactosyl-Sepharose (85 μl) pre-equilibrated with MEPBS containing 0.02% Tween 20 alone (A) or with 0.1 M lactose (B) or 0.1 M sucrose (C) was added to 70 μl of the corresponding supernatant. GII solution (A, B, or C) was then added to the corresponding lactosyl-Sepharose preparation and incubated for 1 h at 24 °C on a rotator. Tubes were centrifuged, and supernatant and pellets were separated, washed, and electrophoresed as described above.

Protein Electrophoresis and Immunoblotting—SDS-polyacrylamide gel electrophoresis according to Fairbanks *et al.* (11) followed by blotting on nitrocellulose (13) was performed as described previously (14). The quantities of the proteins separated on SDS-gels were determined by spectrophotometric quantitation of Coomassie Blue dye eluted from each band (7). Carbohydrate was detected on Western blots after periodate oxidation and testing for the presence of reactive aldehydes according to the manufacturer of the Glycan Detection Kit (Boehringer Mannheim).

Binding of 37-kDa Protein to Immobilized Proteins—Binding of protein to fractions (10 μg) immobilized on nitrocellulose was performed as described elsewhere (7) using antibody followed by ^{125}I-labeled protein A and autoradiography to detect bound 37-kDa protein. To determine the effects of sugars, 26 μg of expressed protein GII in 130 μl of MEPBS without (control) or with lactose (0.15 M), sucrose (0.15 M), galactose (0.18 M), or glucose (0.18 M) were incubated at 24 °C for 30 min on a rotator. Samples were then diluted with 385 μl of MEPBS without (control) or with each sugar at a final concentration of 0.29 M and added to 0.965 ml of blot buffer with 4% bovine serum albumin and 0.3% β-mercaptoethanol containing the blotted protein on bovine serum al-

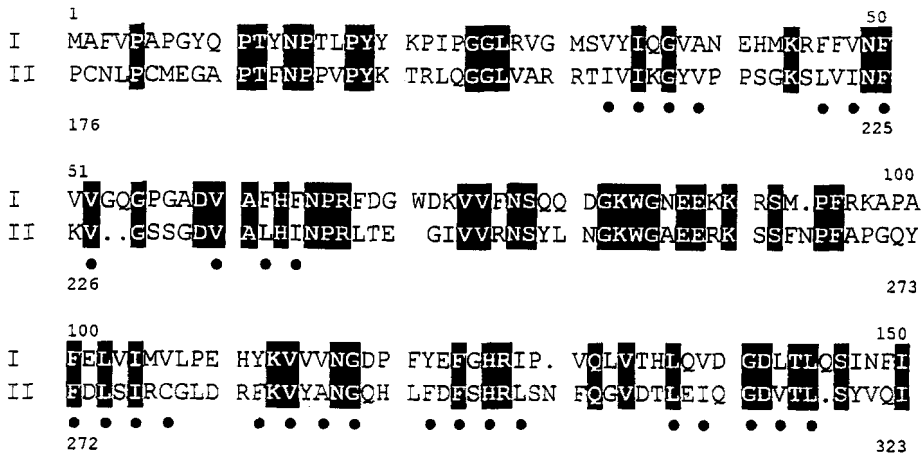

FIG. 3. **The 37-kDa protein contains two homologous domains.** Amino acid sequences were aligned with the "Gap" program of GCG software. Regions with a predicted β-structure and which have an alternating (apolar-*x*) substructure are indicated by • under each of the apolar residues. Seven such strands per domain have been identified. *Boxed* residues are identical in domains I (*I*) and II (*II*).

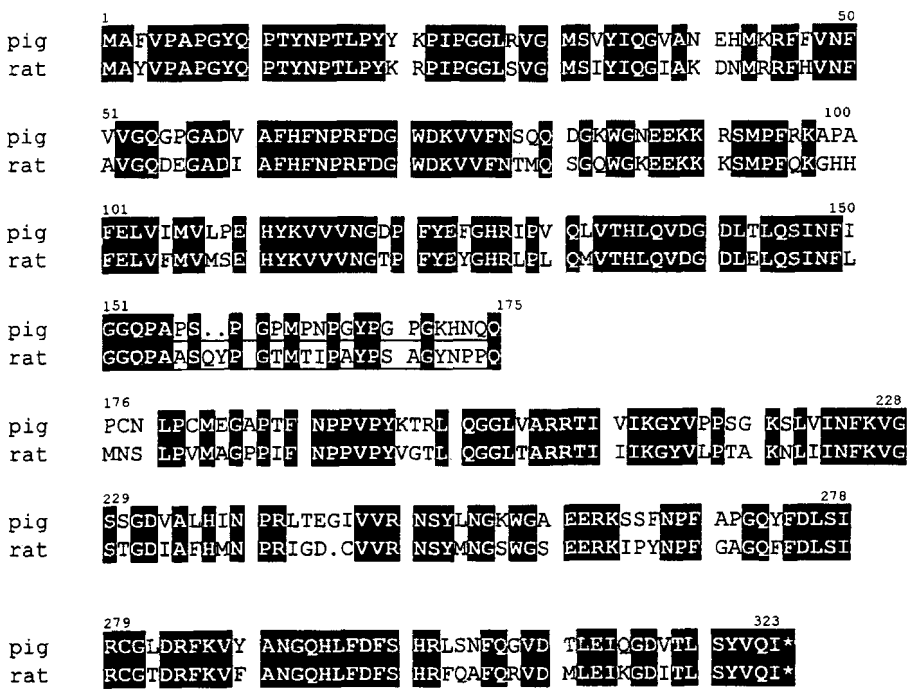

FIG. 4. **Homology of the pig 37-kDa protein and rat L-36 lectin.** Amino acid sequences were aligned with the "Gap" program of GCG software. *Boxed* residues are identical; *underlined* residues are the linking region.

bumin-saturated nitrocellulose (final concentrations: GII, 17 μg/ml; sugar, 0.1 M; and β-mercaptoethanol, 0.2%) and incubated for 2 h at 4 °C with rocking. Nitrocellulose strips were washed and incubated with anti-37-kDa protein antiserum and ^{125}I-labeled protein A for autoradiography as described except that incubation with antibodies was at 4 °C overnight. Inhibition was quantitated by cutting out the 120-kDa band from nitrocellulose and counting on a γ counter.

RESULTS

Sequence of the 37-kDa Protein—Screening of the library with antibody to the 37-kDa protein yielded five clones, of which the largest was about 1.1 kbp. Two were sequenced. Analysis of the 1.1-kbp cDNA showed an open reading frame of 1124 nucleotides encoding a 323-amino acid protein of 35,852 daltons containing two peptide sequences also found in a tryptic digest of the purified protein, with the start codon in a Kozak site (caccATGg) (15) and a polyadenylation signal (nucleotides 1089–1093) and poly(A) tail in the 3′-untranslated region (Fig. 1). The amino acid composition was similar to that determined experimentally from the purified protein (Table I). The size of the mRNA by Northern blot analysis of pig tongue RNA with ^{32}P-labeled 1.1-kbp cDNA probe was about 1.2 kbp (Fig. 2).

Structure of the 37-kDa Lectin—The pig protein consists of two major homologous domains of about 150 amino acids (domain I, residues 1–150, 17,046 daltons; domain II, residues 176–323, 16,375 daltons) (Fig. 3) separated by a 25-residue linking region rich in proline and glycine (residues 151–175, 2468 daltons) (Fig. 1). Since the two similar domains of approximately 16–17 kDa at the N- and C-terminal regions are 39% identical, their tertiary conformations will necessarily be essentially identical. With the Robson and co-workers (16) and the Chou-Fasman (17) predictive techniques, each domain is expected to be rich in β-structure but almost devoid of α-helix. The predicted β-strands are independently recognizable from inspection of the sequence alone, since each shows a regular alternation of apolar and other residues (Fig. 3). Consequently when the seven β-strands hydrogen bond to one another to form what is expected to be a largely antiparallel, right-handed twisted β-sheet, one of its faces will be almost totally apolar in nature. This would be directed inward and away from the aqueous environment. The β-sheets in each domain may fold independently into distinct β-barrel-like entities or, alternatively, may each contribute half of a much larger β-barrel conformation. Both possibilities seem equally likely. The antiparallel

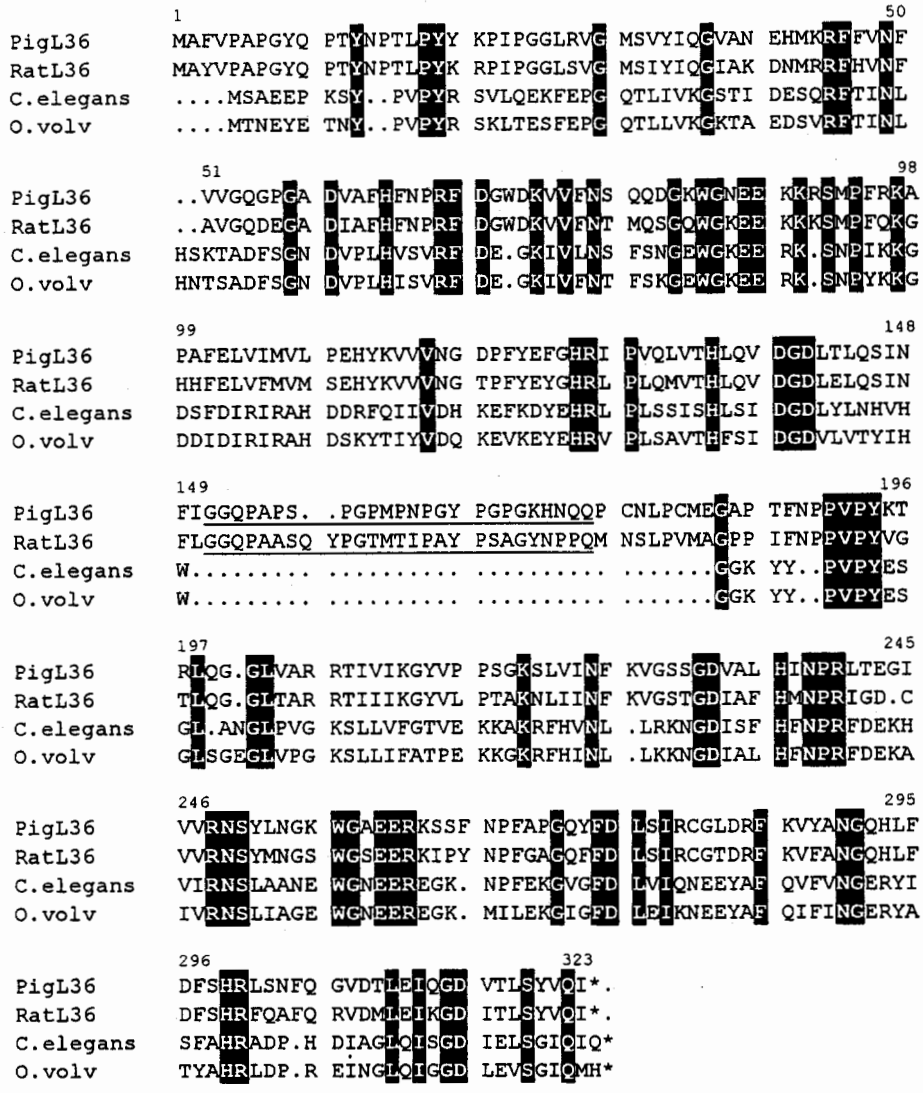

FIG. 5. **Homology of pig L-36, rat intestinal L-36, and lectins from *C. elegans* and *O. volvulus*.** Amino acid sequences were aligned by sequential use of the "Lineup," "Pileup," and "Pretty" programs of GCG software. Identical residues are *boxed*, and *underlined* residues are the linking domain.

nature of much if not all of the β-sheet is deduced from the short lengths of sequence, generally with predicted β-turn structures, that connect consecutive β-strands. Although there are one or two longer loop sequences that could permit parallel β-strands to occur, these appear to be exceptions to the general rule. The right handed nature of the twist of the β-sheet is expected on the basis of numerous observations of similar motifs found in crystalline globular proteins solved at atomic resolution using x-ray diffraction methods.

The two domains are linked by a short sequence (residues 151–175) with unusual characteristics: GGQ$(PX)_6Y(PG)_2$KHNQQ. Not only is the alternation of proline residues unexpected, but the sequence as a whole is rich in apolar residues and sparse in charged ones. There is little doubt that the structure adopted must either be a compact globular one that links domain I to domain II but which is shielded from the water by lying between them or is an internal feature around which both domains are folded. The former possibility seems more likely. It is unequivocal that such an apolar stretch of sequence must be shielded from contact with water. The proline-rich nature of the central domain extends into the N-terminal part of domain II (and by analogy is also present at the N-terminal end of domain I). Proline residues limit conformational flexibility and are found predominantly in structural turns in the tertiary conformation of proteins.

The 37-kDa Protein Is Homologous with Lactose-binding Lectins—When the deduced sequence was compared with known sequences, the highest homology, 71% identity, was with a lactose-binding rat intestinal lectin of unknown function known as L-36 (6) (Fig. 4). The N-terminal 17-kDa domains of the 37-kDa protein and of the rat lectin L-36 were 77% identical, the C-terminal 16-kDa domains were about 72% identical, and the linking domains were 56% identical. The 37-kDa pig protein is presumably the homolog of rat L-36. The pig junction protein showed lesser homologies with a diverse group of proteins of this lectin family (reviewed by Oda *et al.* (6)) from mammalian species including human, rat, mouse, hamster, cow, and pig as well as from chick and other vertebrates (eel) and invertebrates (*C. elegans*).

Rat L-36 was initially described as a soluble 17-kDa fragment from rat intestine which bound to lactose columns and was the first lactose-binding lectin discovered with internal homology (similar large N- and C-terminal domains). The two homologous regions and linking region of pig L-36 correspond well with those of rat L-36. Another reported protein with internal homology of N- and C-terminal domains, the *C. elegans* β-galactoside-binding lectin (18) is 34% identical (59% homologous based on conservative substitutions) with the pig lectin and 36% identical (60% homologous) with the rat intestinal lectin and shows conservation of many of the motifs common to the pig and rat L-36 lectins (Fig. 5). A fourth sequence with two similar domains and homology to the *C. elegans* lectin and to the pig and rat L-36 lectins has been found in *Onchocerca*

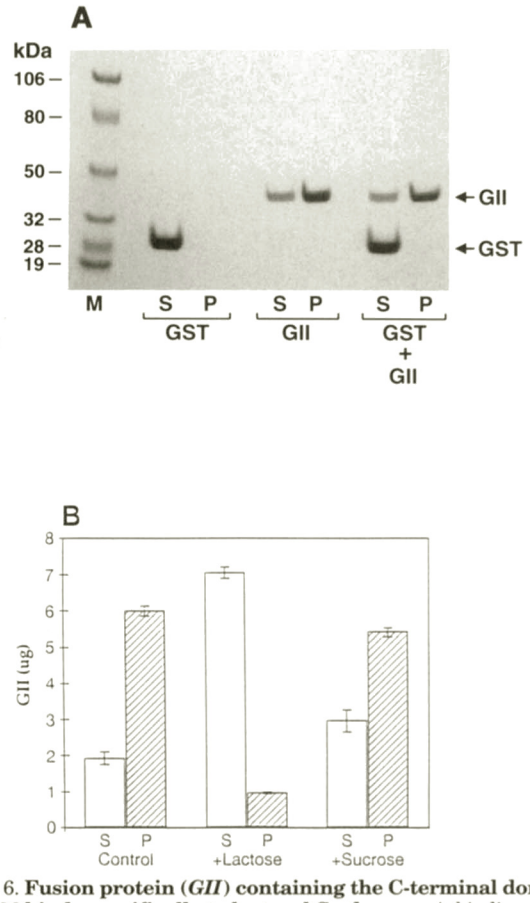

FIG. 6. **Fusion protein (*GII*) containing the C-terminal domain of pig L-36 binds specifically to lactosyl-Sepharose.** *A*, binding of GII to lactosyl-Sepharose is specific for the lectin domain of the fusion protein and does not result from binding of glutathione *S*-transferase (*GST*). GST or the fusion protein of GST and the C-terminal domain of pig L-36 (GII) were incubated with lactosyl-Sepharose as described under "Materials and Methods," and the samples were centrifuged to obtain supernatant (*S*) and pellet (*P*) and analyzed by SDS-polyacrylamide gel electrophoresis. *Lane M*, molecular mass standards; *lanes S*, supernatant was added to Fairbanks sample buffer as described under "Materials and Methods" and electrophoresed; *lanes P*, lactosyl-Sepharose pellet after incubation and washing was extracted as described under "Materials and Methods," and the supernatant was electrophoresed. *GST*, lactosyl-Sepharose incubated with GST. *GII*, lactosyl-Sepharose incubated with GII. *GII + GST*, lactosyl-Sepharose incubated with GII and GST. *B*, binding of GII to lactosyl-Sepharose is inhibited by lactose but not by sucrose. Lactosyl-Sepharose was incubated with GII with or without lactose or sucrose as described under "Materials and Methods," and bound (*P*) and unbound (*S*) fractions were subjected to gel electrophoresis as described in the text. GII was quantitated by extracting bound dye from excised bands as described elsewhere (7). *Control*, lactosyl-Sepharose incubated with GII; *+Lactose*, lactosyl-Sepharose incubated with GII in the presence of 0.1 M lactose; *+Sucrose*, lactosyl-Sepharose incubated with GII in the presence of 0.1 M sucrose. Results are shown as the average ± range of two experiments.

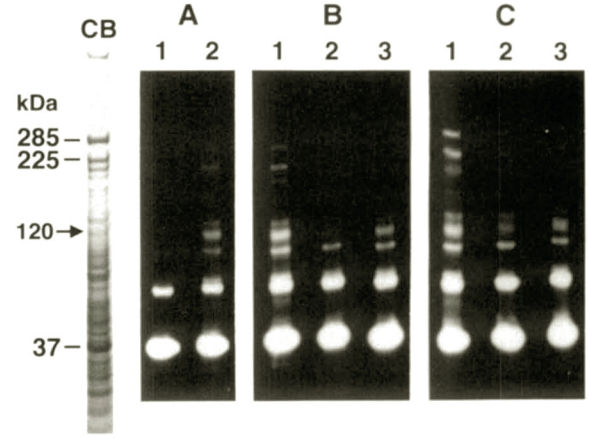

FIG. 7. **Domain II fusion protein (GII) binds to a protein in the junction preparation, and binding is inhibited by lactose.** The citric acid-insoluble fraction of pig tongue epithelium (10 μg) was transferred to nitrocellulose. GII was incubated with nitrocellulose strips containing transfered proteins, and binding of GII was determined as described under "Materials and Methods." *A*, nitrocellulose was incubated with or without GII at 17 μg/ml in the absence of sugars prior to incubation with antibody to the 37-kDa protein. *Lane 1*, control; *lane 2*, with GII. *B*, nitrocellulose was incubated with GII with or without sugars prior to incubation with antibody to the 37-kDa protein. *Lane 1*, GII alone (control); *lane 2*, GII in the presence of 0.1 M lactose; *lane 3*, GII in the presence of 0.1 M sucrose. *C*, as in *B*; *lane 1*, GII alone (control); *lane 2*, GII in the presence of 0.1 M galactose; *lane 3*, GII in the presence of 0.1 M glucose. Radioactivity of the 120-kDa band detected in *B*, counts/min: *lane 1*, 3890; *lane 2*, 700; *lane 3*, 1930; inhibition is 82% for *lane 2* and 50% for *lane 3*. Radioactivity of the 120-kDa band detected in *C*, counts/min: *lane 1*, 3550; *lane 2*, 1510; *lane 3*, 2690; inhibition is 56% for *lane 2* and 24% for *lane 3*. *Arrow*, 120-kDa band. *CB*, molecular mass standards stained with Coomassie Blue.

volvulus[2]; the two invertebrate proteins are 71% identical and 84% similar (conservative substitutions) (Fig. 5). The invertebrate lectins have a truncated or absent linking region.

Properties of Pig L-36—L-36 purified from pig tongue is insoluble and requires high concentrations of urea (approximately 9 M) for extraction from junction preparations (7) or from untreated pig oral epithelium.[3] The C-terminal domain of rat L-36 was initially purified as a soluble protein from rat intestine by Leffler *et al.* (4). Although the lactose-binding lectins are generally soluble, the full rat L-36 sequence expressed in bacteria was insoluble (6), and the expressed pig L-36 clone from the 1.1-kbp cDNA was insoluble as well (data not shown). Since a fragment might be soluble, we expressed the 16-kDa C-terminal domain in bacteria. DNA coding for amino acids 176–323 (domain II) was synthesized by PCR with primers containing *Eco*RI sites (see "Materials and Methods"). The expressed fusion protein was cleaved by thrombin, but free domain II was also insoluble (not shown). The glutathione *S*-transferase fusion protein (designated as GII) was therefore studied.

Initial studies indicated that the soluble fusion protein GII was bound by lactosyl-Sepharose, and binding was produced by the interaction of domain II of the 37-kDa protein and lactosyl-Sepharose, since glutathione *S*-transferase did not bind lactosyl-Sepharose (Fig. 6*A*). The time course of binding was linear from 15 to 60 min (data not shown), and inhibition studies were therefore conducted for 60 min in the presence or absence of sugars. Binding was inhibited by 0.1 M lactose, but not by 0.1 M sucrose. Lactose inhibited binding of the fusion protein to lactosyl-Sepharose by 85%, but sucrose reduced binding by only 10% (Fig. 6*B*).

We reported previously that the purified pig L-36 protein binds to proteins of the junction preparation on Western blots, of which a band of M_r about 120,000 was the most prominent (7). Similarly, using the same methodology as in previous studies, when we incubated the fusion protein (GII) with nitrocellulose strips containing junction proteins from the citric acid-insoluble pellet from pig tongue epithelium (7, 11) transferred to nitrocellulose, the GII protein bound to a protein with M_r = 120,000. To investigate the effect of lactose on the binding of GII to the 120-kDa protein, we incubated GII (20–26 μg/1.5 ml) with nitrocellulose strips containing the citric acid-insoluble proteins (10 μg) with or without sugars, then with antibody to the 37-kDa protein, and subsequently with ^{125}I-labeled protein A (7). GII bound the 120-kDa protein (Fig. 7*A*), and binding was

[2] A. Klion and J. E. Donelson, unpublished results; GenBank™ accession no. U04046.

[3] E. J. O'Keefe and M. L. Chiu, unpublished results.

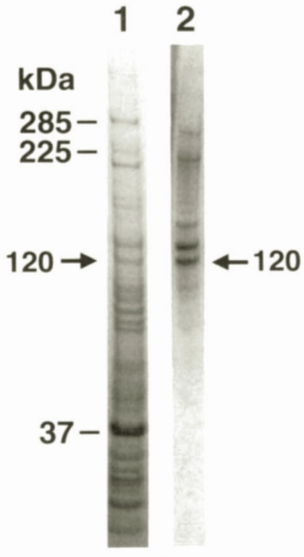

FIG. 8. **The protein which binds the pig L-36 lectin is a glycoprotein.** The citric acid-insoluble fraction from pig tongue epithelium was extracted with buffer containing 4 M urea, and the insoluble fraction was electrophoresed and transferred as described under "Materials and Methods." Proteins transferred to nitrocellulose were tested for carbohydrate content after formation of reactive aldehydes by exposure to periodate as described under "Materials and Methods." *Lane 1*, Coomassie Blue R-250-stained gel (15 μg of protein); *lane 2*, nitrocellulose subjected to glycan test (7.5 μg of protein); *arrow*, 120-kDa protein bound by the pig L-36 lectin.

inhibited by 82% in the presence of 0.1 M lactose (Fig. 7*B*). Sucrose, containing fructose rather than galactose, inhibited binding by about 50%. The monosaccharide galactose inhibited binding by 56% and glucose by 24% (Fig. 7*C*), in agreement with reports indicating that rat (19) and human (20) lectins interact with lactose by interacting not only with galactose but also with glucose. Both the 120-kDa protein and a slightly higher molecular mass protein were found to be glycoproteins (Fig. 8) but pig L-36 did not bind the higher molecular mass band (Fig. 7) or several other prominent proteins, suggesting that binding to the 120-kDa protein was specific.

DISCUSSION

Of the three major categories of the lactose-binding lectins, the smaller M_r 14,000–16,000 lectins, the L-29 family of M_r 29,000–35,000, and the larger L-36 group with a repeating structure of two homologous domains separated by a linking region, only rat L-36 and the homologous *C. elegans* protein have been described in the literature. The L-36 rat intestinal lectin, first discovered and described as a soluble 17-kDa fragment purified from rat intestine, is now known from cDNA sequence analysis (6) to be a 36-kDa protein. The highly homologous pig protein we describe here is the second member of this group found in mammals, and both bind lactose. It is interesting that the invertebrate protein from *C. elegans*, which lacks many of the residues common to the mammalian lactose-binding lectins, also binds lactose; the number and identity of residues critical for this function is not known.

Pig L-36 domains I and II show greater homology with the C-terminal lectin domain of the L-29 group of lectins than with the L-14 subfamily of lectins, 14–16-kDa proteins. Hirabayashi *et al.* (18) noted that domains I and II of the *C. elegans* 32-kDa lectin show homology with the L-14 lectins and resemble a "tandem repeat" of the L-14-type lectins. Oda *et al.* (6) noted that rat L-36 is more closely related to the L-29 subfamily than to the L-14 subfamily by (i) sequence homology and (ii) studies of inhibition of binding to lactose by sugars. Oda *et al.* (6) described variable repeats of proline- and glycine-rich motifs in the N terminus of the L-29 subfamily that in some cases can be cleaved by bacterial collagenase show little homology with lectin domains of the lactose-binding lectins and are reduced or absent in the L-14 and L-36 lectins. Like the rat L-36, the two homologous domains of the pig L-36 protein are more similar to the lectin region of the L-29 group of proteins than to the L-14 lectins, but the differences are not great, and it will be important to make intraspecies comparisons of sequences. Elegant studies of inhibition of binding of lactose by a panel of sugars provide perhaps better evidence for the similarity of L-36 domains I and II with L-29 (6, 19).

These lectins vary in their localization, and although their functions appear to be diverse, detailed functional information is lacking. The smallest lactose-binding lectins, noncovalent dimers with a subunit molecular mass of 14–16 kDa (1, 2, 5) were initially found in the cytoplasm or in the extracellular matrix (2). The murine form was reported to inhibit growth of mouse embryo fibroblasts and to bind to the cell surface, but lactose did not affect binding (21). CBP-35, reported to be localized to the nucleus of proliferating 3T3 cells (22, 23), has been found to be the homolog of L-29 from dog (Madin-Darby canine kidney cells) (24), RL-29 from rat lung (25), rat εBP (IgE binding protein (26), mouse L-29 (27), and human L-29 (28). The IgE binding protein is a cell surface receptor (29). A 67-kDa lactose-binding lectin from bovine chondroblasts was found to have properties of an elastin receptor (30).

Although rat L-36 was isolated as a 17-kDa soluble fragment, the L-36 protein from pig tongue epithelium as well as the cloned and expressed rat L-36 are apparently insoluble, since SDS was required for solubilization (6). Pig L-36 appears to be a cell-cell junction protein, probably an adherens junction, based on colocalization with actin and vinculin, but not with the desmosome protein desmoplakin, and is likely to be located on the cytoplasmic side of the membrane (7). The location in the region of the submembrane skeleton and in a cell-cell junction appears to be unique for a β-galactoside-binding lectin, although localization of the L-36 lectin in rat and *C. elegans* has not been published. It will be important to isolate proteins associated with pig L-36 and to determine whether, as our data suggest, the lectin function may be involved in the assembly of cell-cell adhering junctions, apparently in an intracellular protein-protein interaction.

REFERENCES

1. Barondes, S. H. (1984) *Science* **223,** 1259–1264
2. Barondes S. H. (1986) in *The Lectins. Properties, Functions, and Applications in Biology and Medicine* (Liener, I. E., Sharon, N., Goldstein, I. J., eds) pp. 437–466, Harcourt Brace Jovanovich, New York
3. Barondes, S. H., and Leffler, H. (1987) *Methods Enzymol.* **138,** 510–515
4. Leffler, H., Masiarz, F. R., and Barondes, S. H. (1989) *Biochemistry* **28,** 9222–9229
5. Drickamer, K. (1988) *J. Biol. Chem.* **263,** 9557–9560
6. Oda, Y., Herrmann, J., Gitt, M. A., Turck, C. W., Burlingame, A. L., Barondes, S. H., and Leffler, H. (1993) *J. Biol. Chem.* **268,** 5929–5939
7. Chiu, M. L., Jones, J. C. R., and O'Keefe, E. J. (1992) *J. Cell Biol.* **119,** 1689–1700
8. Ausubel, F. M. (1994) in *Current Protocols in Molecular Biology* (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., eds) John Wiley & Sons, New York
9. Hunter, W. M., and Greenwood, F. C. (1962) *Nature* **194,** 495–496
10. Levi, G., and Teichberg, V. I. (1981) *J. Biol. Chem.* **256,** 5735–5740
11. Fairbanks, G., Steck, T. L., and Wallach, D. F. H. (1971) *Biochemistry* **10,** 2606–2617
12. O'Keefe, E. J., Erickson, H. P., and Bennett V. (1989) *J. Biol. Chem.* **264,** 8310–8318
13. Towbin, H., Staehelin, T., and Gordon, J. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76,** 4350–4354
14. O'Keefe, E. J., Woodley, D., Castillo, G., Russell, N., and Payne, R. E., Jr. (1984) *J. Invest. Dermatol.* **82,** 150–154
15. Kozak, M. (1986) *Cell* **44,** 283–292
16. Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120

17. Chou, P. Y., and Fasman, G. D. (1974) *Biochemistry* **13,** 222–245
18. Hirabayashi J., Satoh, M., and Kasai, K. (1992) *J. Biol. Chem.* **267,** 15485–15490
19. Leffler, H., and Barondes, S. H. (1986) *J. Biol. Chem.* **261,** 10119–10126
20. Sparrow, C. P., Leffler, H., and Barondes, S. H. (1987) *J. Biol. Chem.* **262,** 7383–7390
21. Wells, V., and Mallucci, L. (1991) *Cell* **64,** 91–97
22. Moutsatsos, I. K., Wade, M., Schindler, M., and Wang, J. L. (1987) *Proc. Natl. Acad. Sci U. S. A.* **84,** 6452–6456
23. Laing, J. G., and Wang, J. L. (1988) *Biochemistry* **27,** 5329–5334
24. Herrmann, J., Turck, C. W., Atchison, R. E., Huflejt, M. E., Poulter, L., Gitt, M. A., Burlingame, A. L., Barondes, S. H., and Leffler, H. (1993) *J. Biol. Chem.* **268,** 26704–26711
25. Cerra, R. F., Gitt, M. A., and Barondes, S. H. (1985) *J. Biol. Chem.* **260,** 10474–10477
26. Albrandt, K., Orida, N. K., and Liu, F.-T. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 6859–6863
27. Cherayil, B. J., Weiner, S. J., and Pillai, S. (1989) *J. Exp. Med.* **170,** 1959–1972
28. Robertson, M. W., Albrandt, K., Keller, D., and Liu, F.-T. (1990) *Biochemistry* **29,** 8093–8100
29. Wollenberg, A., de la Salle, H., Hanau, D., Liu, F.-T., and Bieber, T. (1993) *J. Exp. Med.* **178,** 777–785
30. Hinek, A., Wrenn, D. S., Mecham, R. P., and Barondes, S. H.(1988) *Science* **239,** 1539–1541
31. Program Manual for the GCG Package, version 7 (1991) Genetics Computer Group, 575 Science Dr., Madison WI 53711