Eur. J. Biochem. *254*, 217−222 (1998)
© FEBS 1998

Sialic-acid-binding lectin from the slug *Limax flavus*

Cloning, expression of the polypeptide, and tissue localization

Sumiko KURACHI [1], Zhiwei SONG [2], Masaru TAKAGAKI [2], Qing YANG [2], Harry C. WINTER [2], Kotoku KURACHI [1] and Irwin J. GOLDSTEIN [2]

[1] Department of Human Genetics, The University of Michigan, Ann Arbor, USA
[2] Department of Biological Chemistry, The University of Michigan, Ann Arbor, USA

(Received 22 December 1997/16 March 1998) − EJB 97 1798/2

A cDNA library of *Limax flavus* was constructed and screened for sialic-acid-specific lectins. Complementary DNA clones were categorized into seven groups corresponding to closely related but different sequences. Group 1 clones contained an ORF encoding 199 amino acids including a sequence identical to the partial amino acid sequence obtained from the lectin protein. Within its 1074-bp 3′ untranslated region, ten closely related 60-bp sequence repeats were found. Group 2 clones contained an ORF encoding a polypeptide chain of the same number of amino acid residues, with 89.1% overall identity to that of the group 1 and eight 60-bp repeat sequences in the 3′ untranslated region. The remaining groups of clones contained ORF with highly similar full or partial sequences, with or without 60 bp repeats in the 3′ untranslated region. The large number of closely related but different cDNA clones obtained indicated that the slug sialic-acid-specific lectin gene is a member of a multigene family. The lectin amino acid sequence showed significant similarity with the fibrinogen domain of human tenascin-C, with a human C-type serum lectin, and with pig ficolin. Immunostaining analysis of slug tissue for the lectin indicated that it is present primarily on the epidermal surface and in mucous glands. Recombinant slug lectin protein lacking the 20-amino-acid N-terminal signal sequence produced in a bacterial expression system from a group-1 clone accumulated as aggregates in inclusion bodies, suggesting that large-scale production of the active agglutinin may be possible.

Keywords: lectin; sialic acid; *Limax flavus*.

Among sialic-acid-binding lectins of animal origin, the slug (*Limax flavus*) lectin is one of the most useful [1−3]. A homodimer (2×22 kDa [2]), *L. flavus* agglutinin recognizes terminal *N*-acetylneuraminic acid in α2-3-, α2-6-, or α2-8 linkages. A comprehensive study of its binding specificity [3−5] indicated a marked preference for the α-ketosidic configuration at C2, an obligate *N*-acetamido group at C5 (a free amino, *N*-formyl, or *N*-glycolyl group does not bind to the lectin [4]), and tolerance for a substituted hydroxyl group at C4 (e.g. *O*-acetyl or *O*-benzyl). Additionally, although the 4-deoxy derivative binds to *L. flavus* agglutinin, the *epi* derivative, in which the C4 hydroxyl group is in an axial configuration, represents a destabilizing factor [4].

L. flavus agglutinin generates precipitin-like curves with laminin, fetuin and BSA conjugates of sialooligosaccharides [3, 4]. It also reacts with bovine and ovine submaxillary mucins and Tamm-Horsfall protein [1, 3−5]. In an earlier paper, we noted that a cyanogen-bromide-generated peptide (*L. flavus* agglutinin is blocked at its N-terminus) shared similarity with tenascin and fibrinogen [4].

In this paper, we report the cloning of a series of slug-lectin-related cDNAs, tissue-specific localization of *L. flavus* agglutinin in slugs, and production of recombinant slug lectin protein. The similarity of the lectin with other biologically important proteins is discussed.

Correspondence to I. J. Goldstein, Department of Biological Chemistry, The University of Michigan, Ann Arbor, MI 48109-0606, USA

Abbreviations. FITC, fluorescein isothiocyanate; UTR, untranslated region.

Note. The nucleotide sequence data presented here have been submitted to the GenBank database and are available under accession numbers AF060449 (clone 24), AF060450 (clone 11), and AF060451 (clone 6).

MATERIALS AND METHODS

Materials. Restriction and DNA-modification enzymes were obtained from Life Technologies and Pharmacia. [α-^{35}S]dATP, [α-^{32}P]dCTP, rapid hybridization buffer and Megaprime DNA-labeling system were obtained from Amersham. FastTrack mRNA-isolation kit and cDNA library construct Librarian II with pcDNA II were purchased from Invitrogen. The DNA-sequencing kit was obtained from United States Biochemical. Maximum-strength nylon filters were obtained from Schleicher & Schüll. Oligonucleotide primers were synthesized at the DNA core facility on this campus. High-resolution-extended-range-sequencing solution was obtained from National Diagnostics. PCR reagents for constructing the expression vector were obtained from Boehringer-Mannheim. Midi-prep kits for plasmid DNA isolation were from Qiagen. The expression plasmid pET 3a and *Escherichia coli* BL21 (DE 3) were from Stratagene. Polyclonal antiserum to slug lectin was obtained by injecting New Zealand White rabbits subdermally with 50−100 μg native purified slug lectin in Freund's complete adjuvant, followed by 3−4 boosts at 2-week intervals of the lectin in incomplete adjuvant.

Preparation of hybridization probes for library screening. A unique DNA hybridization probe (219 bp) for screening the cDNA library was prepared by PCR using slug tissue geno-

mic DNA as the template. The genomic DNA was isolated from fresh slug tissue according to Blin and Stafford [6] with minor modifications. Forward and reverse PCR primers (5′-GTIGAITTIACITT$^{T}_{C}$AA$^{T}_{C}$AA-3′ and 5′-TTITACCACCAIGCICC$^{A}_{G}$TG-3′, where I represents inosine) were prepared according to partial internal amino acid sequences VEFTFNN and HGAWWYK, respectively, obtained from purified lectin protein ([4], unpublished data). Inosine was used at degenerate codon sites [7]. PCR was performed for 30 cycles of 93 °C for 30 s, 55 °C for 1 min and 72 °C for 1 min. PCR products were subjected to 1.5 % agarose gel electrophoresis and the expected DNA fragments (219 bp) were eluted using the Unidirectional Electroeluter Model UEA (Eastman). The fragments were subjected to sequencing to confirm the sequence of the hybridization probes.

Construction of cDNA library. Slugs were frozen in liquid nitrogen, and the gut was removed under frozen conditions. The frozen, eviscerated body of the slug (1.9 g) was macerated using a mortar and pestle. The mRNA was extracted directly from the tissue sample using the FastTrack mRNA-isolation kit following the manufacturer's instructions. A cDNA library was constructed using a Librarian II kit with phagemid pcDNA II. An aliquot (4.5 µg) of poly(A)-rich RNAs prepared from the slug tissue was reverse transcribed using an oligo(dT) primer linked with a *Not*I linker (Librarian II kit). *Bst*XI linker was ligated to the blunt-ended cDNA, and digested by *Not*I. The double-stranded cDNA fragment generated carried *Bst*XI and *Not*I sticky ends at the 5′ and 3′ ends, respectively. The double-stranded cDNAs were subjected to size selection utilizing gel electrophoresis, and cDNA bands with apparent sizes of 0.5 kb or larger were isolated. These fragments were used for the construction of a cDNA Library with the expression vector pcDNA II.

Screening of library. The cDNA library was screened using the colony-hybridization method [8] with the 219-bp probe described above. The probe was labeled to 2×10^{9} cpm/µg by means of the Megaprime DNA-labeling system with the same primers that were used to generate the 219-bp fragments. After the positive colonies were isolated and purified, phagemid DNAs were prepared by a modified method of the alkaline-extraction procedure [9].

DNA sequencing and database analysis. A series of synthetic oligonucleotides were made as sequencing primers and dideoxynucleotide sequencing was performed using Sequenase Version 2.0 (United States Biochem) according to the manufacturer's directions. The insert DNA fragment was sequenced at least twice by sequencing both strands or the same strand repeatedly, confirming the fidelity of sequence. For the repeats and poly(A) region of the sequence, an automated DNA sequencer (Applied Biosystems, model 373A) at the DNA sequencing core facility on this campus was used. The DNA sequence data were analyzed using the VMS Alpha operating system at the General Clinical Research Center of this campus running the Wisconsin package version 9.0 by Genetics Computer Group.

Slug tissue immunofluorescence staining. Live adult slugs were rinsed several times with distilled water and quickly frozen in liquid N_2. Tissue sections of 8 µm were prepared and placed on poly(lysine)-coated slides. For staining, the tissue sections were incubated with 1% BSA in 10 mM NaP_i, 150 mM NaCl, pH 7.2 ($NaCl/P_i$), for 30 min at room temperature to block the non-specific interactions. After removal of the blocking buffer, rabbit antiserum raised against *L. flavus* agglutinin was diluted 1:100 in blocking buffer, applied to the tissue section and incubated for 2 h. The tissues were then rinsed five times with $NaCl/P_i$. Fluorescein-isothiocyanate (FITC)-labeled goat anti-rabbit antibody was added and incubated for 1 h at room temperature. Alternatively, tissue sections were stained directly, after blocking, with FITC-labeled fetuin, to detect the location of the active lectin. Staining patterns were observed under a fluorescent microscope.

Expression of *L. flavus* agglutinin in *E. coli*. An *L. flavus* agglutinin expression vector was constructed using cDNA clone 11, which has a predicted amino acid sequence identical to that of the native lectin. A PCR-amplified fragment, 560 bp in size, containing the coding region was prepared by means of 5′ and 3′ primers containing *Nde*I and *Bam*HI sites, respectively. The generated fragment was blunt ended and ligated to pUC 18 at the *Sma*I site. Plasmid clones were sequenced, and the *Bam*HI/*Nde*I insert fragment generated was cloned into pET 3a. *E. coli* BL21 (DE3) cells were transformed with the recombinant plasmid (clone S12) and *L. flavus* agglutinin expression was induced with 0.4 mM isopropylthio-*β*-D-galactoside.

Isolation and purification of recombinant *L. flavus* agglutinin. BL21(DE3) cells transformed with the plasmids were harvested by centrifugation at $10000\times g$, and the pelleted cells were washed with 100 mM NaCl, 10 mM Tris/HCl, pH 9.0, 1 mM EDTA (buffer A). Washed cells were suspended in 20 vol. buffer A and sonicated on ice using a Branson Model 250 Sonifier at maximum power for 2 min (30% duty cycle) three times. The sonicates were centrifuged at $12000\times g$ 15 min, and the pellets, containing inclusion bodies, were retained for protein analysis and attempted solubilization and renaturation. Supernatants were retained for agglutination activity assays.

Western blot analysis. SDS/PAGE was performed on 15% gels (2.4% cross-linked) at pH 8.6. The protein bands were transferred onto Immobilon-P or nitrocellulose membranes at pH 10, which were blocked with 5% dry milk or 1.5% BSA overnight, and incubated with anti-(*L. flavus* agglutinin) rabbit serum (1:1000 dilution) for at least 2 h at 4 °C. After washing with buffer at pH 7.5, membranes were incubated with goat anti-rabbit serum conjugated to alkaline phosphatase for 1 h at room temperature, washed with 0.1 M Tris/HCl, 0.1 M NaCl, 50 mM $MgCl_2$, pH 9.5, and developed with 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium.

Hemagglutinin assay. Serially diluted (in buffer A) lectin solutions in 'V'-well microtiter plates were mixed with an equal volume of red blood cell suspension (3% packed cell volume), and agglutination was observed after 1 h. Fresh or formaldehyde-stabilized human cells, generally type A, were used.

RESULTS AND DISCUSSION

A cDNA library of *L. flavus* composed of 4.6×10^{6} independent clones was constructed. By screening the library (72 000 colonies) with the *L. flavus* agglutinin-specific probe, 28 strongly hybridizing clones with insert sizes ranging from 0.5 kb to 1.8 kb were identified and given serial clone numbers. Based on the identity of the predicted amino acid sequences of the ORF, these *L. flavus* agglutinin clones were grouped into seven subgroups. cDNA clones of groups 1, 2 and 3, which consisted of 6, 11 and 2 clones, respectively, were the most complete and were characterized extensively, while clones in the other groups were partial and only analyzed briefly to determine their relationship to those of groups 1, 2, and 3.

In group 1, clones 9 and 11 were identical and the insert was 1708 bp in length, composed of a 34 bp 5′ untranslated region (UTR), a 597-bp ORF encoding 199 amino acid residues, a stop codon, and a 1074-bp 3′ UTR (Fig. 1). The predicted amino acid sequence of the ORF contained a stretch of 47 amino acids identical in sequence to an internal part of the *L. flavus* agglutinin protein sequence that we reported previously [4]. Clones of this group contained ten repeats of a 60-bp sequence in the middle

```
                                   M   S   V   L   F   L   A   A   S   F   L   L   L   A   S   F   E   L
CCAGAGTTGACATCACAAAACTACTCCACT GATC ATG TCC GTT TTG TTC CTC GCT GCG TCG TTC CTG CTG TTG GCC AGC TTT GAA CTT     88

 V   A   A   Q   G   C   P   R   Q   D   P   G   D   W   I   V   I   Q   R   R   L   N   A   D   V   N   F
GTT GCT GCG CAA GGT TGC CCG CGA CAA GAT CCT GGG GAT TGG ATT GTC ATC CAG AGG CGA CTC AAC GCA GAC GTG AAC TTT    169
                 50
 Y   R   G   W   A   D   Y   K   H   G   F   G   D   L   R   C   N   F   W   L   G   N   E   K   I   H   Q
TAC CGT GGC TGG GCC GAT TAC AAA CAT GGG TTC GGA GAC CTC AGG TGC AAC TTC TGG CTT GGG AAT GAA AAG ATT CAT CAA    250

 L   T   S   H   G   R   Y   K   L   R   V   E   F   T   F   N   N   K   S   Y   F   A   E   Y   S   T   F
CTA ACT TCT CAT GGT CGG TAC AAG TTG AGG GTG GAG TTC ACC TTT AAC AAT AAA AGC TAC TTT GCT GAG TAT TCT ACG TTC    331
100
 K   I   L   G   E   A   D   K   Y   K   L   Q   V   G   G   Y   S   G   N   A   G   D   A   L   T   F   H
AAG ATA TTA GGG GAG GCA GAC AAA TAC AAG CTT CAG GTT GGT GGC TAC TCG GGG AAC GCT GGA GAC GCT TTG ACT TTT CAC    412
                                                                                     150
 N   G   M   A   F   S   T   N   D   R   D   N   D   A   D   S   I   D   C   A   K   V   Y   H   G   A   W
AAT GGG ATG GCG TTC AGC ACT AAC GAC CGG GAC AAC GAT GCT GAC TCC ATT GAT TGC GCT AAG GTG TAC CAC GGG GCT TGG    493

 W   Y   K   T   C   H   E   S   N   L   N   G   K   W   G   S   K   K   Y   G   E   G   L   N   W   K   A
TGG TAC AAG ACC TGC CAC GAA AGC AAC CTG AAC GGA AAG TGG GGC TCA AAG AAA TAC GGC GAG GGC TTA AAT TGG AAA GCT    574
                                                                        199
 K   T   T   F   T   A   T   A   T   S   S   L   L   K   I   K   A   L   K  Stop
AAA ACA ACC TTT ACT GCG ACT GCT ACT TCC TCG TTG TTG AAA ATA AAA GCC CTA AAA TAA  CGACAG CGAACATCGC CTAGACTGAA    660
GCACACGAAA GTTGTTCTCT ACTTGCTCTA GTTGGTCTCA GTAAAAAACT AAAATTAAAC TAATTTTCAA CTCTCAATCA CAAAGAAAGA CTAAGGCAAA    760
CCCACTGGTC CACGCCAGCT CCACTGCGGG GAAACATCTG CATAGAAAGA CTAAGGCAAA CCCATTGGTC CACGCCAGCT CCACTGCGGG GAAACATCTG    860
TATAGAAAGA CTAAGGCAAA CCCACTGGTC CACACCAGCT CCACTGCGGG GAAACATCTG TATAGAAAGA CTAAGGCAAA CCCACTGGTC CACGCCAGCT    960
CCACTGCGGG GAAACATCTG CATAGAAAGA CTAAGGCAAA CCCATTGGTC CACGCCAGCT CCACTGCGGG GAAACATCTG TATAGAAAGA CTAAGGCAAA   1060
CCCACTGGTC CACGCCAGCT CCACTGCGGG GAAACATCTG CATAGAAAGA CTAAGGCAAA CCCACTGGTC CACGCCAGCT CCACTGCGGG GAAACATCTG   1160
TATAGAAAGA CTAAGGCAAA CCCACTGGTC CACGCCAGCT CCACTGCGGG GAAACATCTG TATAGAAAGA CTAAGGTAAA CCCACTGGTC CACGCCAGCT   1260
CCACTGCGGG GAAACATCTG TATAGAAAGA CTAAGGCAAA CCCACTGGTC CACGCCAGCT CCGCTGCGGG GAAACATCTG TATAGAAAGA TGCAAAAAGT   1360
AAAAAGTAAA GCGAGTTACA TGTAAACGAG AATCAATTAC AGATATATGC CATGTATTTC AGGTAAACGA AATGAAGATG AAGAAAACGA CTCTCTTTTG   1460
GCCACNGTAG AAACAAGTTT TTATTGTGTG TGGTATTTTA GTCAGGTATA TTATAGTTTC CAGTCCATGT TTCTCACACA CGATAATCTC ATTTCCAGTC   1560
CCTGTTTCTC ACACATGATA ATCTCATTTC CATTCTTTAC ATTACACACA TGAAATGTGA CAATTACCAA TAAAAGTGAG CTTCAGATTA AAAAAAAAAA   1660
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAA                                                           1708
```

Fig. 1. Nucleotide and deduced amino acid sequences of group-1 *L. flavus* agglutinin cDNA. Nucleotide numbering is shown on the right, and the deduced amino acid sequence is shown by single letter codes with residue numbering. Repeat sequences in the 3′ UTR are underlined.

A

```
                                                                  tcacaa
Repeat 1  AGAAAGACTA AGGCAAACCC ACTGGTCCAC GCCAGCTCCA CTGCGGGGAA ACATCTGCAT
Repeat 2  AGAAAGACTA AGGCAAACCC ATTGGTCCAC GCCAGCTCCA CTGCGGGGAA ACATCTGTAT
Repeat 3  AGAAAGACTA AGGCAAACCC ACTGGTCCAC ACCAGCTCCA CTGCGGGGAA ACATCTGTAT
Repeat 4  AGAAAGACTA AGGCAAACCC ACTGGTCCAC GCCAGCTCCA CTGCGGGGAA ACATCTGCAT
Repeat 5  AGAAAGACTA AGGCAAACCC ATTGGTCCAC GCCAGCTCCA CTGCGGGGAA ACATCTGTAT
Repeat 6  AGAAAGACTA AGGCAAACCC ACTGGTCCAC GCCAGCTCCA CTGCGGGGAA ACATCTGCAT
Repeat 7  AGAAAGACTA AGGCAAACCC ACTGGTCCAC GCCAGCTCCA CTGCGGGGAA ACATCTGTAT
Repeat 8  AGAAAGACTA AGGCAAACCC ACTGGTCCAC GCCAGCTCCA CTGCGGGGAA ACATCTGTAT
Repeat 9  AGAAAGACTA AGGTAAACCC ACTGGTCCAC GCCAGCTCCA CTGCGGGGAA ACATCTGTAT
Repeat 10 AGAAAGACTA AGGCAAACCC ACTGGTCCAC GCCAGCTCCG CTGCGGGGAA ACATCTGTAT
          AGAAAGAtgc aaaaa
```

B

```
Repeat 1  accaaga--A AGGCACACCC ACTGGTCCAC GCCAGCTCCA CTGCGGGGAA ACATCTGTAT
Repeat 2  AGAAAGACTA AGGCACACCC ACTGGTCCAC GCCAGCTCCA CTGCGGGGAA ACATCTGTAT
Repeat 3  AGAAAGACTA AGGCACACCC ACTGGTCCAC GCCAGCTCCA CTGCGGGGAA ACATCTGTAT
Repeat 4  AGAAAGACTA AGGCACACCC ACTGGTCCAC GCCAGCTCCA CTGCGGGGAA ACATCTGTAT
Repeat 5  AGAAAGACTA AGGCACACCC ACTGGTCCAC GCCAGCTCCA CTGCGGGGAA ACATCTGTAT
Repeat 6  AGAAAGACTA AGGCACACCC ACTGGTCCAC GCCAGCTCCA CTGCGGGGAA ACATCTGTAT
Repeat 7  AGAAAGACTA AGGCACACCC ACTGGTCCAC GCCAGCTCCA CTGCGGGGAA ACATCTGTAT
Repeat 8  AGAAAGACTA AGGCACACCC ACTGGTCCAC GCCAGCTCCA CTGCGGGGAA ACATCTGTAT
          AGAAAGAtgc aaaaa
```

Fig. 2. Nucleotide sequences of 60-bp repeats of *L. flavus* agglutinin cDNA clones 11 (A) and 24 (B). Nucleotide residues referred to in the text as showing unique combinations are underlined.

of its 3′ UTR (Fig. 2A). These repeats had a set of distinct combinations of single base differences at three positions, namely A...C...C, A...C...T, A...T...C or A...T...T, suggesting their unique duplication during the evolution process. In group 2, clones 23 and 24 were identical, with an insert length of 1.8 kb composed of a 35-bp 5′ UTR, a 597-bp ORF encoding 199 amino acids with 80% identity to that of group 1, a stop codon, and a 1177-bp 3′ UTR containing eight 60-bp repeats. The 60-bp repeat sequences in this group were identical (Fig. 2B), with the same three-position combination of C...C...T, which is different from those of group 1 (Fig. 2B). Group 3 consisted of clones 3 and 6 with partial inserts of 0.8 kb and 0.85 kb, respectively. Clones in this group contained ORF encoding an amino acid sequence with the highest similarity to that of group 2 (89% identity), and slightly lower similarity (83%) to that of group 1 (Fig. 3). This group contained 60-bp repeats in the 3′ UTR, but they were only partial. Group 4 consisted of three clones (8, 15 and 18) with varied insert sizes (0.6, 1.6 and 1.6 kb, respectively). Clone 18 contained only a partial insert sequence with the 5′ half truncated, but its partial ORF coding region had 89% amino acid sequence identity with that of group 1. The predicted amino acid sequence of this group showed an eight-amino-acid extension (IPKDSSRA) at the C-terminal end. The inserts of group 5 clones (clones 2, 13 and 14 with insert sizes of 1−1.1 kb) were found to be unique. These inserts encoded a partial C half lectin sequence, at least 132 amino aicd residues in length. The partial lectin sequence had 84% identity to inserts of groups 1, 2, and 3. Clones in groups 6 and 7 contained only short partial inserts, and no further characterization was made.

A database search showed that the amino acid sequence of *L. flavus* agglutinin predicted from these clones has significant similarity with sequences in human tenascin-C [10], a human serum C-type lectin, P35 [11], and pig ficolin [12] (Fig. 4). The similarities reside in fibrinogen-like domains contained in all of these proteins. Although the overall identity of the *L. flavus* agglutinin sequence with these other proteins is fairly low (35−44%), the middle portion (approximately 150 residues) of the *L. flavus* agglutinin sequence (clone 11) exhibits much higher similarity. For example, in the 27-amino-acid stretch of *L. flavus* agglutinin residues 107−133 (tenascin-C residues 2099−2125), the identity is 63% (74% similarity including conservative substitutions), and the nucleotide-sequence identity is 62%. On the other hand, the 22-residue N-terminal signal peptide and the 32-residue C terminus of *L. flavus* agglutinin are completely dissimilar with the corresponding sequences of fibrinogen-like domains.

A short region (≈160 bp in length) in the 3′ UTR adjacent to the poly(A) tail of *L. flavus* agglutinin clone 11 also showed

```
SASL 11  MSVLFLAASF  LLLASFELVA  AQGCPRQDPG  DWIVIQRRLN  ADVNFYRGWA
SASL 24  ----------  ---T------  -A---P----  ----------  ---D------
SASL  6  ----------  ---T---V--  -K---P----  ----------  ---D---D-V

SASL 11  DYKHGFGDLR  CNFWLGNEKI  HQLTSHGRYK  LRVEFTFNNK  SYFAEYSTFK
SASL 24  ----------  --Y-------  ----------  ----------  ----------
SASL  6  ----------  --Y-------  ----------  ----V----R  ----------

SASL 11  ILGEADKYKL  QVGGYSGNAG  DALTFHNGMA  FSTNDRDNDA  DSIDCAKVYH
SASL 24  V---------  E---------  ---AYQ----  -T-K--H--K  Y-KN--IQ--
SASL  6  ----------  E--------A  -H-AI-----  -T-K------  ----------

SASL 11  GAWWYKTCHE  SNLNGKWGSK  KYGEGLNWKA  KTTFTATATS  SLLKIKALK
SASL 24  ------A--N  -----L--TT  EF-Q--I-RQ  T--H--SP--  TVM---SID
SASL  6  ----------  -----L--TT  KF-Q--S--Q  T--H--SP--  TVM---S D
```

Fig. 3. Comparison of deduced amino acid sequences of the ORF of *L. flavus* agglutinin cDNAs of groups 1 (11), 2 (24) and 3 (6). Identical sequences are shown by dashes.

```
LFA clone 11     1  MSVLFLAASF  LLLASFELVA  AQGCPRQDPG  DWIVIQRRLN  ADVNFYRGWA    50
HS-lectin      111  RGHFLSGWHT  IYLPDCRPLT  VLCDMDT-G-  G-T-F---VD  GS-D---D--   160
Pig Ficolin α  121  RGHILSGWHT  IYLPDCQPLT  VLCDMDT-G-  G-T-F---SD  GS-D---D--   170
Tenascin C    1992  DTTSGLYTIY  LNGDKAQALE  VFCDMTS-G-  G---FL--K-  GRE---QN-K  2041

LFA clone 11    51  DYKHGFGDLR  CNF..WLGNEKI  HQLTSHGRYK  LRVEFT.FNNK  SYFAEYSTFK   100
HS-lectin      161  T--Q---SRL  GE-..----DN-  -A--AQ-TSE  ---DLVD-EDN  YQ--K-RS--   211
Pig Ficolin α  171  A--R---SQL  GE-..----DH-  -A--AQ-TNE  ---DLVD-EGN  HQ--K-RS-Q   221
Tenascin C    2042  A-AA----R-  EE-LH---LDNL  NKI-AQ-Q-E  ---DLRD.HGE  TA--V-DK-S  2093

LFA clone 11   101  ILGEADKYKL  QVGGY.SGNAG  DALTFHNGMA  FSTNDRDNDA  DSIDCAKVYH   150
HS-lectin      212  VAD--E--N-  VL-AFVE-S--  -S-----NQS  ---K-Q---L  NTGN--VMFQ   262
Pig Ficolin α  222  VAD--E--M-  VL-AFVE----  -S--S--NSL  -T-K-Q---Q  YASN--VL-Q   272
Tenascin C    2094  VGDAKTR---  K-E--.--T--  -SMAY---RS  ---F-K-T-S  AITN--LS-K  2143

LFA clone 11   151  GAWWYKTCHE  SNLNGKWGSK  KYG...EGLNWKA  KTTFTATATS  SLLKIKALK*   199
HS-lectin      263  ------N--V  -----RYLRG  TH-SFAN-I---S  GKGYNYSYKV  -EM-VRPA*    313
Pig Ficolin α  273  -----NS--V  -----RYLGG  SH-SFAN-V--SS  GKGYNYSYKV  -EM-FR-T*    323
Tenascin C    2144  --F--RN--R  V--M-RY-DN  NHS...Q-V--FH  WKGHEHSIQF  AEM-LRPSN  2193
```

Fig. 4. Comparison of the deduced amino acid sequence of *L. flavus* agglutinin cDNA clone 11 with those of fibrinogen domains of human serum c-type lectin P35, pig ficolin *α*, and tenascin-C. Identical sequences are shown by dashes; spaces introduced for maximum alignment are shown by dots.

partial similarity with the mast-cell function-associated antigen in the 3′ UTR.

Through the database search, the 60-bp tandem repeat sequences in the 3′ UTR were shown to have significant similarity (71−75%) with parts of other sequences, such as the mRNA for *Oryctolagus cuniculus* eukaryotic initiation factor 2Bε [13], and the human β 2 gene for β tubulin in the reverse orientation [14]. The repeat sequences of group-1 clones showed only a few nucleotide differences, and its tandem arrangement in the cluster without any spacer sequences suggests a precise genetic mechanism for multiplication of the repeats. The eight identical repeat sequences of group 2 are uniquely different from those of group 1, indicating that this group of genes took an independent evolutionary path from that of group 1 after divergence from the primordial gene.

Possible functions of these unique repeats in the 3′ UTR are not known. A tandem repeat of this size and sequence is different from other repeats, such as Alu and Line 1 repeats, which are much longer and dispersed throughout the genome [15, 16]. Much shorter repeats, such as AT or GC in the UTR or trinucleotides such as GAC, GAA or CCG in coding regions, are known to be involved in gene regulation or in genetically determined pathologies, respectively [17−19].

The presence of multiple closely related but different cDNA with dispersed nucleotide differences indicates that the slug sialic-acid-specific lectin is encoded by a family of closely related genes, which apparently evolved from a common ancestral gene. Alternate splicing of mRNA from a single gene could not account for the dispersed differences seen in the ORF of the cDNA clones. The origin of the 60-bp repeats in the course of evolution of the multigene family cannot be ascertained, but their variation in sequence and number suggest that they have evolved with the multigene family during at least part of its evolution. Some plant lectins have been shown to have multigene families [20], suggesting that this phenomenon is a common evolutionary event generating many structurally and/or functionally related but distinct proteins. A possible common function incumbent upon glycosyltransferases, glycohydrolases and lectins is recognition of specific extended oligosaccharide structures.

Localization of *L. flavus* agglutinin in slug tissue. A representative cross-section of slug integument immunolabeled with anti-(*L. flavus* agglutinin) serum and FITC-labeled secondary antibody is shown in Fig. 5. The skin of slugs consists of a single layer of epithelial cells, in which unicellular mucous glands are embedded that penetrate into connective tissue lying above a muscle layer [21]. Fluorescent labeling occurred only on the epidermal surface and in the mucous glands, indicating that the lectin is associated with the mucous secretions that lubricate and protect the soft skin. The active lectin was also labeled directly via its sialic-acid-binding site using FITC-labelled fetuin. In this case, the epidermal surface was uniformly labeled by the fluo-

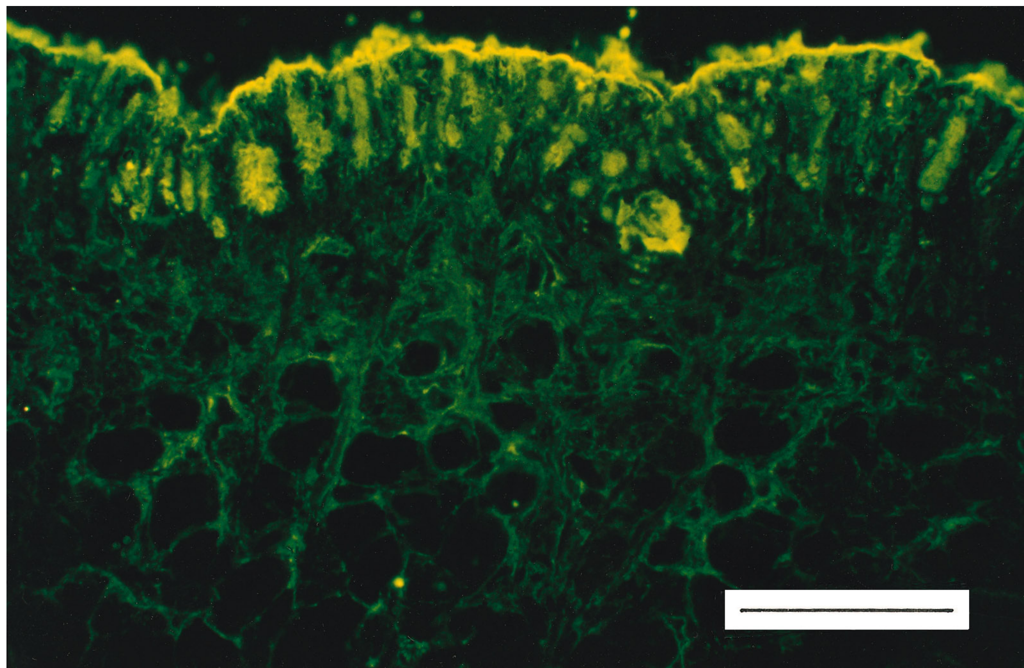

Fig. 5. Localization of *L. flavus* agglutinin in *L. flavus* tissue stained with FITC-labeled anti-(*L. flavus* agglutinin) Ig. Scale bar = 200 μM.

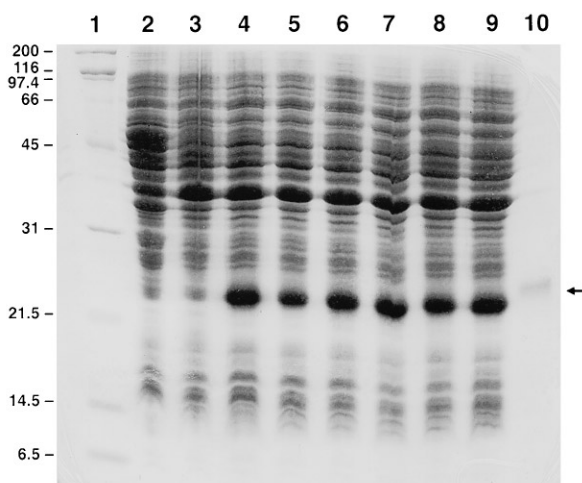

Fig. 6. Expression of slug lectin gene in pET3a-transformed *E. coli* BL21 (DE3). Lane 1, SDS/PAGE standards; lane 2, pET3a/BL21 (DE3) (untransformed); lane 3, pET3a-11/BL21 (DE3) before induction; lanes 4−9, pET3a-11/BL21 (DE3) − pET3a-16/BL21 (DE3) after induction (0.4 mM isopropylthiogalactoside, 3 h, 37 °C); lane 10, native slug lectin (arrowed).

rescent conjugate, although mucous glands were not observed in these sections (data not shown). Apparently, the mucous glands are restricted to limited areas of the slug body, as has been reported for other species of slugs [22].

Expression of recombinant *L. flavus* agglutinin in *E. coli*. The plasmid used for *L. flavus* agglutinin expression encodes the 179-amino-acid *L. flavus* agglutinin peptide (clone 11) lacking the hydrophobic signal-sequence region. The vector contained a coding sequence starting with Met-Gln instead of 5-oxoproline present in the native lectin. One of nine clones (S12) in PET 3a was used to transfect *E. coli* BL21 (DE3) for *L. flavus* agglutinin expression. Transfected cells produced *L. flavus* agglutinin of

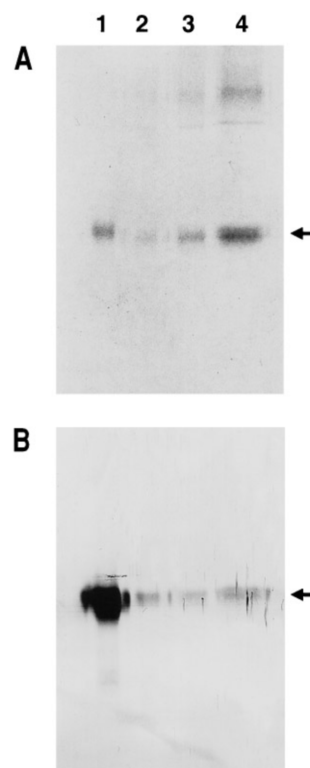

Fig. 7. 15% SDS PAGE (A) and western blot analysis (B) of solubilized inclusion bodies from induced recombinant *E. coli* cells. Lane 1, 2 μg native slug lectin; lanes 2−4, 0.5, 1.0 and 2.0 μg recombinant protein.

approximately 22 kDa into inclusion bodies, as estimated by SDS/PAGE under reducing conditions (Fig. 6). Amino acid sequencing of this protein showed it to contain the N-terminal (H_2N-MQGXPRQDPGDWIVIQRRLNADVNF) and C-terminal sequences after tryptophan-specific cleavage with 2-(2′-nitro-

phenylsulfenyl)-3-methyl-3-bromoindolenine (KAKTTFTATA-TSSLLKIKALK-COOH) that matched the predicted amino acid sequence (Cys, denoted as X, was not detected by the sequencing procedure). Western blot analysis with *L. flavus* agglutinin-specific antibodies also detected this 22-kDa band (Fig. 7). Thus, the product of the *L. flavus* agglutinin construct is expressed in the *E. coli* expression system, and accumulates as an insoluble aggregate in the inclusion bodies.

Recombinant *L. flavus* agglutinin produced in the bacterial system failed to give reproducibly significant soluble agglutination activity. Attempts to form active lectin by solubilization of inclusion bodies in 5 M guanidine hydrochloride and slow dilution of the denaturant in 'refolding' buffer [23] also could not reproducibly recover agglutinin activity nor protein that bound to an appropriate affinity matrix. These attempts usually led to reaggregation of the protein during subsequent dialysis against Tris/NaCl buffer. Native *L. flavus* agglutinin protein, on the other hand, remained soluble and recovered most of its activity under the same conditions of denaturation and refolding.

To address this problem and the potential function of the 60-bp repeats in the 3′ UTR, expression studies with a Chinese hamster ovary cell line lacking sialic acid are in progress. The present work lays the foundation for further studies on the structure/function relationship of the gene and protein, and the biological role of *L. flavus* agglutinin.

This work was supported by National Institutes of Health grant GM29470 to I. J. G. We thank Dr Albert Chu of E.-Y. Laboratories, San Mateo CA, for the generous gift of live slugs.

REFERENCES

1. Miller, R. L. (1982) *J. Invertebr. Pathol. 39*, 210−214.
2. Miller, R. L., Collawn, J. F. Jr & Fish, W. W. (1982) *J. Biol. Chem. 257*, 7574−7580.
3. Knibbs, R. N., Goldstein, I. J., Ratcliffe, R. M. & Shibuya, N. (1991) *J. Biol. Chem. 266*, 83−88.
4. Knibbs, R. N., Osborne, S. E., Glick, G. D. & Goldstein, I. J. (1993) *J. Biol. Chem. 268*, 18524−18531.
5. Fischer, E. & Brossmer, R. (1995) *Glycoconj. J. 12*, 707−713.
6. Blin, N. & Stafford, D. W. (1976) *Nucleic Acids Res. 3*, 2303−2308.
7. Takahashi, Y., Kato, K., Hayashizaki, Y., Wakabayashi, T., Ohtsuka, E., Matsuki, S., Ikehara, M. & Matsubara, K. (1985) *Proc. Natl Acad. Sci. USA 82*, 1931−1935.
8. Hanahan, D. & Meselson, M. (1983) *Methods Enzymol. 100*, 333−342.
9. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res. 7*, 1513−1523.
10. Nies, D. E., Hemesath, T. J., Kim, J.-H., Gulcher, J. R. & Stefansson, K. (1991) *J. Biol. Chem. 266*, 2818−2823.
11. Matsushita, M., Endo, Y., Taira, S., Sato, Y., Fujita, T., Ichikawa, N., Nakata, M. & Mizuochi, T. (1996) *J. Biol. Chem. 271*, 2448−2454.
12. Ichijo, H., Hellman, U., Wernstedt, C., Gonez, L. J., Claesson-Welsh, L., Heldon, C.-H. & Miyazono, K. (1993) *J. Biol. Chem. 268*, 14505−14513.
13. Asuru, A. I., Mellor, H., Thomas, S. B., Lu, Y., Chen, J.-J., Crosby, J. S., Hartson, S. D., Kimball, S. R., Jefferson, L. S. & Matts, R. L. (1996) *Biochim. Biophys. Acta 1307*, 309−317.
14. Lee, M. G.-S., Loomis, C. & Cowan, N. J. (1984) *Nucleic Acids Res. 12*, 5823−5836.
15. Kariya, Y., Kato, K., Hayashizaki, Y., Himeno, S., Tarui, S. & Matsubara, K. (1987) *Gene* (*Amst.*) *53*, 1−10.
16. Dombroski, B. A., Scott, A. F. & Kazazian, H. H. Jr (1993) *Proc. Natl Acad. Sci. USA 90*, 6513−6517.
17. Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., Lawton, M., Trottier, Y., Lehrach, H., Davies, S. W. & Bates, G. P. (1996) *Cell 87*, 493−506.
18. Campuzano, V., Montermini, L., Molto, M. D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., Zara, F., Canizares, J., Koutnikova, H., Bidichandani, S. I., Gellera, C., Brice, A., Trouillas, P., De Michele, G., Filla, S., De Frutos, R., Palau, F., Patel, P. I., Di Donato, S., Mandel, J.-L., Cocozza, S., Koenig, M. & Pandolfo, M. (1996) *Science 271*, 1423−1427.
19. Gacy, A. M., Goellner, G., Juranic, N., Macura, S. & McMurray, C. T. (1995) *Cell 81*, 533−540.
20. Van Damme, E. J. M., Smeets, K., Torrekens, S., Van Leuven, F. & Peumans, W. J. (1994) *Eur. J. Biochem. 221*, 769−777.
21. South, A. (1992) *Terrestiral slugs: biology, ecology, control*, Chapman & Hall, London.
22. Arcadi, J. A. (1963) *Ann. N. Y. Acad. Sci. 106*, 451−455.
23. De Bernardez-Clark, E. & Georgiou, G. (1990) *ACS Symp. Ser. 470*, 1−20.