Novel Galactose-binding Proteins in Annelida

CHARACTERIZATION OF 29-kDa TANDEM REPEAT-TYPE LECTINS FROM THE EARTHWORM LUMBRICUS TERRESTRIS*

Jun Hirabayashi, Samir Kumar Duttaš, and Ken-ichi Kasai

From the Department of Biological Chemistry, Faculty of Pharmaceutical Sciences, Teikyo University, Sagamiko, Kanagawa 226-0012, Japan.

© 1998 by The American Society for Biochemistry and Molecular Biology, Inc.
Printed in U.S.A.

Novel type lectins were found in the phylum Annelida, i.e., in the earthworm, tubifex, leech, and lugworm. The lectins (29–31 kDa) were extracted from the worms without the use of detergent and purified by affinity chromatography on asialofetuin-agarose. On the basis of the partial primary structures of the earthworm Lumbricus terrestris 29-kDa lectin (EW29), degenerate primers were synthesized for use in the reverse transcriptase-polymerase chain reaction. An amplified 155-base pair fragment was used to screen a cDNA library. Four types of full-length clones were obtained, all of which encoded 260 amino acids, but which were found to differ at 29 nucleotide positions. Since three of them resulted in non-silent substitutions, EW29 mRNA was considered to be a mixture of at least three distinct polynucleotides encoding the following proteins: Ala44-Gly197-Ile213 (clone 5), Gly44-Gln197-Ile213 (clone 7), and Ala44-His197-Ile213 (clones 8 and 9; different at the nucleotide level, but encoding an identical polypeptide). Genomic polymerase chain reaction using DNA from a single worm revealed that the single worm already had four sets of cDNAs. The EW29 protein showed two features. First, the lectin was composed of two homologous domains (14,500 Da) showing 27% identity with each other. When each of the domains was separately expressed in Escherichia coli, the C-terminal domain was found to bind to asialofetuin-agarose as strongly as the whole protein, whereas the N-terminal domain did not bind and only retardation was observed. EW29 was found to exist as a monomer under non-denaturing conditions. It had significant hemagglutinating activity, which was inhibited by a wide range of galactose-containing saccharides. Second, EW29 contained multiple short conserved motifs, “Gly-X-XX-Gln-X-Trp.” Similar motifs have been found in many carbohydrate-recognition proteins from an extensive variety of organisms, e.g., plant lectin ricin B-chain and Clostridium botulinum 33-kDa hemagglutinin. Therefore, these carbohydrate-recognition proteins appear to form a protein superfamily.

Galactose-binding proteins (lectins) represent a distinguished group among lectins because statistics have shown that more than 60% of the lectins reported thus far are galactose-specific (1). A possible explanation for such a preference is that galactose was selected as an important “recognition” saccharide especially in higher organisms. In fact, galactose is used more frequently in higher organisms like mammals, whereas glucose and mannose are recognized by microorganisms such as bacteria and fungi. From a glycochemical viewpoint, galactose has a nature inherently distinct from that of glucose, mannose, or fructose. The latter three monosaccharides are related by “Lobry de Bruyn-Alberta van Ekenstein transformation” (2). It is well established that in N-linked oligosaccharide biosynthesis galactose is incorporated at later stages after removal of glucosyl and mannosyl residues from the common precursor Glc3Man9GlcNAc2 (3, 4). These observations led one of the authors to present a hypothesis on the origin of elementary hexoses that galactose is a “latecomer” saccharide relative to glucose and mannose (5). Galactose is exposed at outermost spaces of cells unless it is masked by sialic acids, so that it is easily recognized by various communication molecules by homophilic carbohydrate-carbohydrate (6) or heterophilic carbohydrate-protein interactions (7).

Among galactose-specific lectins, galectins are unique in that all of the members belonging to this family are galactose-specific (8, 9). In general, they are soluble and require no metal ion for their activity. Galectins bind most preferentially to lactosamine-containing saccharides of both glycoproteins and glycolipids. Although galectins had long been believed to occur in vertebrates only, they proved to be distributed in much lower organisms as well (10–15). Although the biological roles of galectins are not fully understood, they are supposed to play multiple roles basic to multicellular organisms, such as in development, differentiation, tumorigenesis, metastasis, apoptosis, etc. In this regard, if galectins are essential for these biological processes, they would be expected to exist in all animal species. However, there have been only few reports of galectins other than in two animal phyla representing deuterostomes and protostomes, i.e., Vertebrata and Nematoda, respectively. On the other hand, if this is not the case, a possibility emerges that some other galactose-binding lectins compensate for the absence of galectins in such animal phyla.

We have undertaken screening of galectin-like proteins from the phylum Annelida, by employing the same purification strategy as that for galectins, i.e., lactose-specific extraction in the absence of detergent and metal ion, and affinity chromatography on asialofetuin-agarose. As a result, 29–31 kDa lectins were purified from four annelids as follows: earthworm, tubifex, leech (both belonging to the class Oligochaeta), lugworm (Polychaeta), and leech (Hirudinea). These proteins immunologically cross-reacted with one another and had similar bio-
chemical properties as galecins with respect to specificity, solubility, and metal independence. Detailed structural analysis including cDNA cloning was carried out for 29-kDa lectin from the earthworm Lumbricus terrestris. As a result, the earthworm 29-kDa lectin (designated hereafter EW29) proved to be a “tandem repeat”-type lectin, which consists of two tandemly repeated homologous domains (14.5 kDa). Contrary to our expectation, it showed no sequence homology to the known galecins. However, it showed resemblance to some carbohydrate-releasing proteins such as ricin B-chain, C. botulinum hemagglutinin, etc. It also contained conserved multiple repeats of short motifs.

EXPERIMENTAL PROCEDURES

Worms—Earthworms (Allolobophora japonica and L. terrestris), tubifex (Tubifex hattai), and lugworm (Neanthes japonica) were purchased from local fishing shops. Leeches (Glossiphonia complanata) were harvested from shallow irrigation water in Toyama Prefecture, Japan. The worms were rinsed with cold PBS, weighed, and processed immediately for lectin purification.

Purification of Annelid Lectins—Lectins were extracted from worms by essentially the same procedure previously described for galecins (16, 17); briefly, worms were disrupted by homogenization in a Polytron (Kinematica) in 5 volumes of cold McCartney ethanol, 2 mM EDTA, 20 mM sodium phosphate, pH 7.2, 150 mM NaCl. After the bulk of soluble proteins had been removed by centrifugation (15,000 rpm, 4 °C, 25 min), the lectins were specifically extracted from the precipitate (p-p-1) with EDTA-MEPBS containing 20 mM lactose by shaking for 30 min at 4 °C. After centrifugation as above, the obtained extract (sup-2) was extensively dialyzed to remove lactose and then applied to a column of asialofetuin-agarose (bed volume, 10 ml) pre-equilibrated with 25 mM EDTA-PBS, 25 mM sodium phosphate, pH 7.2, 150 mM NaCl. After washing of the column with EDTA-MEPBS (400 ml), bound protein was eluted with the same buffer containing 20 mM lactose.

Protein concentration was determined by use of the Bio-Rad Protein Assay Dye Reagent. Protein-containing fractions were subjected to sodium dodecyl sulfate-polyacrylamide gel (14%) electrophoresis (SDS-PAGE) under reducing conditions, unless otherwise stated. Protein was visualized with a Wako Silver Stain Kit.

Separation of EW29 by High Performance Gel-permeation Chromatography—The main component (29 kDa) of the affinity purified lectin fraction from earthworms (EW29) was further purified by high performance gel-permeation chromatography on a TSK-G2000SWXL column (7.5 × 300 mm) equilibrated with 10 mM Tris-HCl, pH 9.0, and 0.02% NaN₃. The cell was collected and lysed by sonication, and recombinant protein was adsorbed to asialofetuin-agarose (5 ml of 1% (w/v) bovine serum albumin in saline, and 25 ml of EDTA-PBS) both in the presence and absence of 20 mM lactose at a flow rate of 0.5 ml/min. Protein elution was monitored by absorbance at 280 nm. For the subsequent hemagglutination assay, lactose was removed by extensive dialysis against EDTA-PBS.

Hemagglutination Assay—Basically, the conventional assay system described by Nowicki et al. (19) was used. 25 µl of serially diluted samples was mixed in each well of a 96-well microtiter V plate with 25 µl of EDTA-PBS, 25 µl of 1% (w/v) bovine serum albumin in saline, and 25 µl of trypanized rabbit erythrocytes. After the plate had stood at room temperature for 1 h, “dot” (no agglutination) or “mat” (agglutination) formation was judged. For assessment of the inhibitory effect of various mono- and oligosaccharides, 25 µl of maximally diluted lectin solution that gave mat formation and 25 µl of serially diluted inhibitor saccharides were used in place of serially diluted samples and EDTA-PBS, respectively. Minimum concentrations that gave negative dot formation under the above conditions were defined as I₅₀.

Effect of Saccharides on Extraction of Earthworm Lectin—The effect of various saccharides on lectin extraction was investigated as described previously (12); earthworm lectin was extracted from small portions of ppt-1 (equivalent to 1-g weight of the worm) in the presence of a 0.1 M concentration of various sugars. After centrifugation, the supernatant solutions were subjected to SDS-PAGE, followed by Western blotting on a nitrocellulose membrane. The lectin was stained by a conventional double antibody method using anti-tubifex lectin antisem (described below) and horseradish peroxidase-conjugated goat anti-rabbit IgG antiserum (Seikagaku Co., Tokyo). Both antisera were used at a 1,000-fold dilution. For peroxidase detection, POD Immunostain Set or High Sensitive Immunoblotting Kit (both from Wako Chemicals, Tokyo) was used.

Production of Antisera—Antisera were raised in rabbits against either affinity purified tubifex lectin or affinity and gel filtration-purified EW29 by injecting the animals with 1 mg of purified protein in 0.1 ml of Freund’s incomplete adjuvant. After the injection, the produced antisera was evaluated by both dot and Western blotting analyses. The antisera were stored at 4 °C in the presence of 0.02% NaN₃.

Protein Structural Analyses—Affinity purified lectins from earthworm tubifex, lugworm, and leech were further purified by reversed phase chromatography on a TSK-ODS-80TM column (7.5 × 250 mm) and were analyzed by a protein sequence (Applied 477A) as described (17). The peptides were designated “Lys”-1, -2, and -3 and so on, in the order of elution. In addition, in the case of earthworm 29-kDa lectin, reduction and S-carboxymethylation was also performed prior to the protease digestion. Thereseurgethirdpeptidewerepreboxed“Cm-Lys.”

Preparation of Genomic DNA from Earthworms—Genomic DNA was prepared from two earthworm species, A. japonica and L. terrestris, by the conventional phenol-chloroform procedure. Genomic DNA were used for the preparation from A. japonica and a single worm was used for that from L. terrestris to assess the observed gene polymorphism (described below).

cDNA Cloning of EW29—A probe DNA was prepared by means of the polymerase chain reaction (PCR). For amplification of a part of earthworm lectin gene, four sets of convergent oligonucleotide primers were synthesized based on the determined sequence peptides (Lys-4, Lys-7, Lys-8, and Lys-10), designated as Lys-4F/Lys-4R, Lys-7F/Lys-7R, Lys-8F/Lys-8R, and Lys-10F/Lys-10R (see Table IV for detail). PCR was performed by use of all combinations of the above forward (F) and reverse (R) primers and a Takara LA-PCR Kit (94 °C, 30 s; 52 °C, 1 min, 72 °C, 3 min for 35 cycles, then 72 °C, 10 min). Genomic DNA prepared from the earthworm A. japonica was used as a template. The amplified DNA (155 bp) was obtained with primers Lys-4F and Lys-7R was cloned into pCR II (Invitrogen) on the basis of TA-cloning strategy according to the manufacturer’s instructions, and the nucleotide sequence was confirmed with an Applied 373S DNA sequencer.

For cloning of full-length cDNAs, a ZAP cDNA library (kindly provided by Dr. G. Zieve, Düsseldorf, Germany) was screened with the above cDNA fragments labeled with [γ-32P]dCTP (Pharmacia) and an Amersham Megaprome Kit. A standard plaque hybridization was performed for screening 3 × 10⁸ plaques as described by Sambrook et al. (20). Eight positive plaques were obtained and converted into plasmid pBluescript (Stratagene) by an automatic in vivo excision system according to the manufacturer’s instructions. The plasmids were cut with EcoRI and XhoI to confirm the insert size and subjected to sequence analysis and cloning. (anpliing An application, 25 µl was used.)

Production of Recombinant Earthworm Lectins, Wh, Nh, and Ch—Expression plasmids for recombinant EW29 protein (Wh), N-terminal domain (Nh; meaning N-half), and C-terminal domain (Ch) were constructed by a PCR procedure; for the production of Wh, a pair of oligonucleotide primers was designed, i.e. G ATG GCT GGA AGG CCT TTC CTT CTG GTG (Met1-Ala-Gly-Pro-Phe-Leu; designated Wh-F; see Table IV) and CGAGTGGAGT TTA CTC GGA TTC G (antisense, Glu 258-Ser-Ter; Wh-R), and a full-length fragment (794-bp) encoding 260 amino acids was amplified with a full-length plasmid (clone 5) used as a template. The derived fragment was cloned into pCR II in frame with β-galactosidase a-peptide, and the derived plasmid was used to transform Escherichia coli TOP10F (Invitrogen). Positive clones were selected by colony immuno blot analysis using specific antisera against EW29 and horseradish peroxidase-conjugated goat anti-rabbit IgG as described previously (21). The cloned E. coli was proliferated to full growth in 1 liter of LB medium in the presence of antibiotics at 37 °C. The recombinant protein was induced by adding isopropyl-β-thiogalactoside to give a final concentration of 0.2 mM and incubating the culture for 2 h at the same temperature. The cells were collected and lysed by sonication, and recombinant protein was adsorbed to asialofetin-agarose, as described previously (10).

For construction of plasmids for expression in different domains (Nh and Ch), pairs of oligonucleotide primers for an interdomain region were designed, i.e. C ATG ACG CCG AAG TTC TAC TCT ACG (NhEnd) and Lys-Pro-Lys-Pho-Pho-Tyr-Tyr (designated Ch-F; underline denotes introduced mutations) and TTA GAA GAA CCT CGG CTT CAA GTG (NhStart).
Identification of Galactose-binding Lectins in Annelids—
Based on the concept that galectins are bound to insoluble glycoconjugates but are solubilized with a competing sugar, lactose, we applied the same procedure to purify annelid lectins as employed for galectins (10, 16, 17). Since it became evident that annelid lectins are structurally related, subsequence analyses, summarized in Table II, strongly suggest that these annelid lectins are structurally related. In particular, most of the lectins had multiple similar sequences represented by “Gly-X-X-Gly-X-Trp.” Furthermore, some of the sequences containing this short motif showed highest similarity in case of earthworm lectin, these included His-Gly-Glu-Arg-Lys-Asn-Glu-Trp (NH-7), and Asn-Gly-Glu-Pro-Asn-Gln-Ala-Trp (NH-10; conserved residues underlined). This indicates that the annelid lectin contained a significant number of repeats.

Effect of Saccharides on Lectin Extraction—Since it became evident that annelid lectins are structurally related, subsequent studies were focused on the earthworm 29-kDa lectin (EW29), because its purification yield was the best among the species examined (Table I). First, various saccharides were tested for their ability to extract the lectin from the earthworm precipitate (ppt-1) as described under “Experimental Procedures.” As a result, among various simple saccharides, only lactose, lactate, and melibiose were found to be effective. On maintenance of the homogenate at a low temperature, the proportion of the 16-kDa species in Fig. 1B was considerably decreased (data not shown). (ii) Among the four species of annelids, the amount of smaller fragments was highest in the most fragile worm, tubifex (see Fig. 1B). (iii) The sizes of smaller fragments were almost half of those of the main species (29–31 kDa). To confirm the above possibility, we isolated the 16-kDa earthworm protein by reversed phase chromatography and compared its peptide sequences with those of EW29. Since both molecular species gave identical peptide sequences (data not shown), at least some of the smaller fragments were considered to be degradation products of the 29–31-kDa proteins, probably due to cleavage at the connecting region between tandemly repeated domains (described below).

Comparison of Structures of Annelid Lectins—In order to elucidate the structural relationship between the annelid lectins, we first attempted Western blotting analysis by using anti-tubifex lectin antiserum. As shown in Fig. 2, all of the lectins purified from the other annelid species were also positive, together with the smaller fragments described above. The degree of cross-reactivities with leech (Hirudinea) and lugworm (Polychaeta) proteins was relatively weak, probably reflecting phylogenetic distance between tubifex and earthworm (both belonging to the same class, Oligochaeta). When crude lectin extracts (sup-2) were used instead of purified fractions, only high molecular species (29–31 kDa) were observed. This is consistent with the above assumption that smaller fragments were digestion products.

For more direct structural comparison, affinity purified lectins from earthworm, tubifex, and leech were further purified by reversed phase chromatography on a TSK-TMS250 (data not shown) and digested with lysine-specific Achromobacter protease I. The derived peptides were separated on a TSK QDS-80TM column and were analyzed by a protein sequencer. For earthworm lectin, we also carried out reduction and S-carboxymethylation to identify cysteine residues. The results of peptide analyses, summarized in Table II, strongly suggest that these annelid lectins are structurally related. In particular, most of the lectins had multiple similar sequences represented by “Gly-X-X-Gln-X-Trp.” Moreover, some of the sequences containing this short motif showed highest similarity; in case of earthworm lectin, these included His-Gly-Gln-Asp-Ala-Gly-Gln-Trp (Lys-4), His-Gly-Gly-Thr-Asp-Gln-Gln-Trp (Lys-7), and Asn-Gly-Gly-Pro-Asp-Gln-Ala-Trp (Lys-10; conserved residues underlined). This indicates that the annelid lectin contained a significant number of repeats.

TABLE I
Summary of purification yields of galactose-specific lectins from four annelid species

| Worm class | Starting material, wet weight | Purified on ASF-agarose* | Yield |
|------------|------------------------------|--------------------------|-------|
|            | g mg mg/g worm               |                          |       |
| Tubifex (Oligochaeta) | 27 0.58 21 |                         |       |
| Earthworm (Oligochaeta) | 30 1.50 50 |                         |       |
| Lugworm (Polychaeta) | 10 0.17 17 |                         |       |
| Leech (Hirudinea) | 4.4 0.17 59 |                         |       |

* ASF-agarose is asialofetuin-agarose.

(antisense, HisLys-Leu-Lys-Pro-Lys-Phe-Phe-Ter; Nh-R). For production of Nh and Ch, primer sets Wh-F/Nh-R and Ch-F/Wh-R were used, respectively. All other procedures were the same as those employed for Wh.

Single Worm Genomic PCR—In order to assess a cause for the polymorphism found in the earthworm lectin cDNAs (described under “Results”), genomic DNA was prepared from a single worm, and genomic PCR was performed with primers Wh-F and Wh-R to amplify a full-length region. After cloning of the derived fragments (0.8 kilobase pairs) into pCRII and transformation of E. coli TOP10F, 9 independent genomic clones were selected and their nucleotide sequences were compared with one another and to those derived from the cDNA cloning.

RESULTS

Identification of Galactose-binding Lectins in Annelids—
Based on the concept that galectins are bound to insoluble glycoconjugates but are solubilized with a competing sugar, lactose, we applied the same procedure to purify annelid lectins as employed for galectins (10, 16, 17) as follows: (i) galactose-specific extraction in the absence of metal ions, and (ii) affinity chromatography on asialofetuin-agarose. A typical result of affinity chromatography on asialofetuin-agarose is shown for earthworm lectin in Fig. 1A, and the results of SDS-PAGE for all four annelid species, i.e. earthworm, tubifex, leech and lugworm, are shown in Fig. 1B. Major molecular species observed after lactose elution from the column were the following: 29 kDa (earthworm), 31 kDa (tubifex), 30 kDa (leech), and 30 kDa (lugworm). The observed sizes were also similar to those reported for either mammalian chimeria-type (29–35 kDa) (22) or nematode tandem repeat-type galectins (32 kDa; see Refs. 10, 11, and 13). Since these lectins were obtained in the absence of metal ion, they were considered not to be C-type lectins. The fact that no detergent was necessary for the extraction also excluded the possibility that the annelid lectins are membrane-bound proteins. The yields of purified lectins from the four annelid species are summarized in Table I.

Beside major molecular species (29–31 kDa), earthworm, tubifex, and leech preparations also showed the presence of multiple smaller fragments of 15–17 kDa (Fig. 1B). However, the following observations suggest that these fragments were derived by proteolysis of the 29–31-kDa species. (i) When purification of earthworm lectin was processed as rapidly as possible (e.g. within 2 days) and special care was taken for the
the other hand, glucose, α-methylmannoside, L-fucose, maltose, and sucrose had no effect even at the highest concentration used (100 mM; data not shown). As the next step, concentrations of the effective saccharides were changed (i.e., 1, 10, and 100 mM), and the amounts of extracted lectin were compared (Fig. 3). Lactose and melibiose showed almost the same ability to extract the lectin at 10 and 100 mM, whereas galactose showed slightly poorer ability. It is notable that EW29 was extractable with not only lactose, but also galactose and melibiose (Galβ1,6Glc), a linkage isomer of lactose (Galβ1,4Glc). Such a feature has never been observed for galectins.

**Oligomeric Structure and Hemagglutinating Activity of EW29**—High performance gel filtration analysis was carried out to estimate the molecular organization of EW29 under non-denaturing conditions. The affinity purified earthworm lectin (a mixture of major 29-kDa and minor 15–17-kDa fragments) was applied to a TSK G2000SWXL column in the presence of 20 mM lactose (Fig. 4A), since in its absence the protein elution was considerably retarded. Fractions showing UV adsorption (280 nm) were subjected to SDS-PAGE analysis (Fig. 4B). By comparing eluting positions of marker proteins, the 29-kDa lectin was considered to exist as a monomer (calculated mass, 30,000 Da). SDS-PAGE analysis showed that the purified fraction was >95% pure and almost free of the 16-kDa species. We used this fraction as highly purified EW29 for the production of specific antisera and for the following hemagglutination test.

Thus purified EW29 agglutinated trypsinized rabbit erythrocytes at the minimum concentration of approximately 10 μg/ml (final concentration in the assay, 2.5 μg/ml). This agglutinating ability means that EW29 has more than one sugar-binding site per 29-kDa polypeptide, as it was proved to exist as a monomer. A panel of saccharides was tested for inhibitory activity toward hemagglutination as described under “Experimental Procedures.” No saccharides lacking galactose showed inhibition. This is consistent with the result of lectin extraction as described above. On the other hand, galactose-containing saccharides showed varied inhibitory effects on hemagglutination, which were expressed as $I_{50}$ in Table III. The results are summarized as follows. (i) As regards monosaccharides, galactose had similar inhibitory activity as its N-acetylated (GalNAc) and α-methylated (α-Me-Gal) derivatives, but its β-glycosides were significantly stronger inhibitors than galactose; i.e., Gal, GalNAc, α-Me-Gal < β-Me-Gal < β-pAP-Gal. (ii) Among disaccharides, melibiose showed the same inhibitory power as lactose, but their N-acetyl derivatives, N-acetyllactosamine (type 2) and Galβ1,6GlcNAc as well as another linkage isomer, Galβ1,3GlcNAc (type 1), were still better inhibitors. Notably, similar inhibitory power was also observed for Galβ1,3GalNAc (T antigen). However, the non-reducing synthetic disaccharide thiodigalactoside (Galβ1,4Gal) was the best inhibitor among those tested. (iii) L-Fucosylation of lactose at either galactose (L-Fucα1,2Galβ1,4Glc) or glucose moiety (Galβ1,4(L-Fucα1,3)Glc) reduced the activity by 1 order.

The above binding feature was common to galectins in some ways but significantly distinct in others (8). They were similar in showing (i) moderate affinity to lactose, (ii) significantly higher affinity to both N-acetyllactosamine (type 2) and its linkage isomer Galβ1,3GlcNAc (type 1), and (iii) the strongest affinity to thiodigalactoside among disaccharides. However, EW29 showed strong affinities for monosaccharides, particularly β-Me-Gal and β-pAP-Gal, whereas galectins never have such a preference. EW29 also showed a wider preference for disaccharides such as melibiose and Galβ1,3GalNAc, to which galectins in general do not bind. However, the latter is exceptionally well recognized by rat galectin-4 (23) and nematode 16-kDa galectin.2 Thus, EW29 is basically galactose-specific but seems to exhibit much wider preference for galactose-containing saccharides than galectins.

**cDNA Cloning of EW29**—Based on the peptide sequences of EW29 (Lys1, Lys3, Lys6, and Lys8), four pairs of degenerate primers were synthesized (for their sequences, see Table IV). PCR was performed by using all combinations of these forward and reverse primers, and genomic DNA was prepared from the earthworm as a template. A significant amplification product (155 bp) was obtained only when the combination of primers, Lys-4F and Lys-7R, was used (data not shown). The amplified fragment was cloned into pCRII, and the nucleotide sequence was determined. The deduced amino acid sequence was found to follow the sequence of Lys-4 (used as a forward primer) with no termination codon in frame and also included the determined sequence of Lys-3, “Asp-Val-Val-His-His-Arg-Asp-Asp-Lys” (data included in Fig. 5). This confirmed the derived fragment is relevant as a PCR product corresponding to EW29.

Next, the above 155-bp PCR fragment was used as a probe to screen a λ-ZAP cDNA library of the earthworm *L. terrestris* (kindly provided by Dr. T. Giebing, Düsseldorf). As a result of screening 3 × 105 plaques, 8 positive clones were obtained (designated clones 1, 2, 3, 5, 7, 8, 9, and 10), and their sequences were determined following conversion into plasmid Bluescript as described under “Experimental Procedures.” We found the following. (i) Clones 5, 7, 8, and 9 were of full-length clones, all of which consisted of 261 codons including ATG (Met) initiation and TAA termination codons. (ii) Clones 1, 3, 5, and 10 contained identical sequences, but their lengths were different (non-full-length clones 1, 3 and 10 started with G48, C76, G16, respectively. Numbers begin at initiation codon A'TG). Clone 2 is almost identical to clone 5, but differed at two positions, i.e., CGG (clone 2)/CGA (clone 5) both encoding Arg at 16th position and GTG (clone 2)/GTT (clone 5) encoding Val. (iii) The remaining clones 7, 8, and 9 were apparently distinct. They differed from clone 5 at 12, 15, and 10 positions, respectively, in the coding region (sequence identities to clone 5: 98.5, 98.1, and 98.7%, respectively). All of these differences were only single nucleotide substitutions, and no deletion or insertion was observed.

---

2 H. Ahmed, et al., unpublished observations.
Earthworm 29-kDa Lectin

Numbers on the first amino acid residues of most earthworm peptides are the position numbers (initial Met as position 1). Lowercase letters represent amino acids for which determination is ambiguous. Prefix Cm represents peptides derived by lysylendopeptidase digestion after EW29 was denatured and S-carboxymethylated. Bold letters denote conserved three amino acid residues in the repetitive motif, GxxGxGxxxG. Underlined residues indicate corresponding nucleotide sequence used as a PCR primer.

| Peptide | Sequence | Comment |
|---------|----------|---------|
| Earthworm | SNTAPGA | Used as a PCR primer |
| Ly6 | D194VHHRNDK | |
| (Mix) | S138ELNGK | |
| Ly8 | ENQRWRK | |
| Ly8\(\text{a}\) | P11AEHGQDAQQWK | |
| Ly8\(\text{b}\) | V144LDIEGQNFAPGSK | |
| Ly8\(\text{c}\) | H156QIETQFDPNPXK | |
| (Mix) | MAPHSGDsHSQRWFGDGNK | |
| Ly8\(\text{d}\) | Y116HGHTNQQWETVHLK | Used as a PCR primer |
| Ly8\(\text{e}\) | INHPxMDIENGQLSRFPQHG | Used as a PCR primer |
| Ly8\(\text{f}\) | hGIAALASANASGTPVwSVK | |
| Ly8\(\text{g}\) | YQPNGGPNQAWTFEFVGGQAQPGYPSV | Used as a PCR primer |
| Cm-Lys1 | WVLNGHND | |
| Cm-Lys2 | PACHTED | |
| Cm-Lys3 | V144LDIEGQNPAPG | =Ly\(\text{a}\) |
| (Mix) | E237AGAHicawk | |
| Cm-Lys4 | V97FD1cRGSk | |
| (Mix) | E106QIETQFDPNPXk | =Ly\(\text{g}\) |
| Cm-Lys5 | Dx106VGAQV | |
| Cm-Lys6 | Y116IGGNYNQ | =Ly\(\text{g}\) |
| Cm-Lys7 | GFPVQ | |
| Tubifex | wK | |
| Ly3 | xxzDKxGENREDGA | |
| Ly3\(\text{a}\) | IEXYwVLxwYIAF | |
| Ly3\(\text{b}\) | NGANNQVFFEYV | |
| (Mix) | ANGDGEQ | |
| Ly8 | VQRREYMTSELMGqa | |
| Ly8\(\text{a}\) | GSANQHYIIEY | |
| Ly8\(\text{b}\) | NVHLQPFPNGGATQQwVVDGTAK | |
| Ly8\(\text{c}\) | NQLWLYQDQGQFIRSLTDFALDGQa | |
| Leech | | |
| Ly8 | MQPYNNGAK | |
| Ly8\(\text{a}\) | HQWRIEGNK | |
| Ly8\(\text{b}\) | REFMIK | |
| Ly8\(\text{c}\) | NQLwYQQSGSLr | |
| (Mix) | GxELxAYxFK | |
| Ly8\(\text{d}\) | NGPNqexTIVYntk | |
| Ly8\(\text{e}\) | NQLwYQQSGGLIRSSLNDFV | |

*These fractions contained a mixture of two peptides.

Nucleotide and deduced amino acid sequences of clone 5, as a representative, is presented in Fig. 5. The sequences of clones 5, 7, 8, and 9 are also available with GenBank™/EMBL/DDBJ accession numbers AB010783, AB010784, AB010785, and AB010786, respectively.

Most of the above differences between clones were found at the third letter of codons (24 cases out of 29; see Table V). Even if they occurred at the first base (4 cases), 3 of them were at the third letter of codons (24 cases out of 29; see Table V). Even if they occurred at the first base (4 cases), 3 of them were at the third letter of codons (24 cases out of 29; see Table V). The values well agreed with those obtained by genomic PCR using DNA prepared from a single worm of *L. terrestris*. The entire coding region was amplified by

**Table II**

Summary of peptide sequences derived by lysylendopeptidase digestion of annelid lectins

The amino acid substitutions were rather conservative. The resulting three polypeptides encoded by these four types of cDNA clones were as follows: Ala\(\text{44}\)-Gln\(\text{197}\)-Ile\(\text{213}\) (encoded by clone 5), Gly\(\text{44}\)-Gln\(\text{197}\)-Val\(\text{213}\) (clone 7), and Ala\(\text{44}\)-His\(\text{197}\)-Ile\(\text{213}\) (both clones 8 and 9). Calculated molecular masses for these variants were 29,006, 28,978, and 29,015 Da, respectively. The values well agreed with those obtained by SDS-PAGE (29,000 Da). The nucleotide sequences did not contain any hydrophobic signal sequence for secretion or membrane integration, or any signal for N-linked oligosaccharide attachment (Asn-X-(Ser/Thr)).
As expected, a 35-kDa protein (a fusion product with b-riders. Earthworm was homogenized with EDTA-MEPBS in the absence of saccharide. After centrifugation, the soluble fraction (sup-1) was then divided into 10 portions, from which lectin was extracted with MEPBS containing one of the following saccharides: 100, 10, 1 mM lactose (lanes 1–3), respectively, 100, 10, 1 mM melibiose (lanes 4–6), and 100, 10, 1 mM galactose (lanes 7–9). Each lane contained 10 μg of protein. Glucose, methyl-D-mannoside, L-fucose, maltose, and sucrose had no effect on lectin extraction (data not shown).

![Image](http://example.com/figure1)

**Fig. 3. Extraction of earthworm lectin with various saccharides.** Earthworm was homogenized with EDTA-MEPBS in the absence of saccharide. After centrifugation, the soluble fraction (sup-1) was then divided into 10 portions, from which lectin was extracted with MEPBS containing one of the following saccharides: 100, 10, 1 mM lactose (lanes 1–3), respectively, 100, 10, 1 mM melibiose (lanes 4–6), and 100, 10, 1 mM galactose (lanes 7–9). Each lane contained 10 μg of protein. Glucose, methyl-D-mannoside, L-fucose, maltose, and sucrose had no effect on lectin extraction (data not shown).

![Image](http://example.com/figure2)

**Fig. 4. Separation of EW29 by high performance gel-permeation chromatography.** A, affinity purified earthworm lectin (50 μg), consisting of 29 kDa and smaller molecular species (15–17 kDa), was applied to a column of TSK G2000SWXL (7.5 × 300 mm) equilibrated with EDTA-PBS containing 40 mM sodium phosphate and 20 mM lactose. Protein was eluted at a flow rate of 0.5 ml/min while monitoring absorbance at 280 nm and was fractionated every 0.5 min from when the first protein elution began (15 min). Marker proteins used for calibration were as follows: bovine serum albumin (66,000 Da), chicken ovalbumin (45,000 Da), bovine carbonic anhydrase B (29,000 Da), and bovine pancreatic ribonuclease A (13,800 Da). **B**, the result of SDS-PAGE for the fractionated proteins. Electrophoresis was carried out under reducing conditions using 14% gel. Protein was stained with silver.

using newly synthesized primers, Wh-F and Wh-R (Table IV), and the derived fragments (0.8 kilobase pairs) were analyzed after cloning into pCRII. As a result, all of the 9 genomic clones analyzed proved to be correlated to one of the four types of cDNA clones as follows: 1, 2, 3, and 3 genomic clones were assigned to cDNA clone 5, 7, 8, and 9, respectively. This result suggests that the bands a, c, and d were C-terminal fragments beginning from Lys, His, and Lys, respectively. Since all these positions were located in the connecting region between N-terminal and C-terminal domains, this result indicates that only the C-terminal domain retained binding activity. On the other hand, the band “b” did not give any significant amino acid peak even when a sufficient amount of protein was applied. It might be attributable to artificial N-terminal cyclization to form pyroglutamate, because many glutamine residues, i.e., Gln, Gln, and Gln, are located in the connecting region.

We further constructed expression plasmids for both individual N-terminal and C-terminal domains of EW29. Two mutant primers, 384F (Ch-5) and 405R (Nh-3’), were synthesized (Table IV). The individual N-half and C-half regions were expressed in E. coli, and their binding properties were tested. The C-terminal domain bound to the column (Fig. 7B) as stably as the whole molecule, whereas the N-terminal domain did not bind (data not shown). However, even in the latter case, elution from the asialofetuin-agarose column was significantly retarded (Fig. 7A). Therefore, we concluded that the C-terminal domain has stronger binding ability than the N-terminal domain, as far as asialofetuin-agarose was used, although the latter retains some affinity.

**DISCUSSION**

When we first chose tubifex (class Oligochaeta) for purification of galactose-binding proteins, multiple protein bands around 15–17 kDa together with a faint 31-kDa band were observed on SDS-PAGE after purification by asialofetuin-agarose chromatography. However, antiserum raised against this mixed fraction detected only the 31-kDa species in the lactose extract of tubifex (data not shown). This implied that the smaller fragments were degradation products generated by proteolysis during purification from the parental 31-kDa protein. This antiserum also detected approximately 30-kDa protein. This antiserum also detected approximately 30-kDa protein.

| Saccharides | IC<sub>50</sub> (μM) |
|-------------|---------------------|
| Monosaccharides |             |
| Galactose    | 0.1–0.2             |
| GalNAc       | 0.1–0.2             |
| α-Me-Gal     | 0.1                 |
| β-Me-Gal     | 0.05                |
| β-p-AP-Gal<sup>a</sup> | 0.02 |
| Disaccharides |                  |
| Lactose (Gal-β1-4Glc) | 0.1          |
| Melibiose (Gal-β1-4Glc) | 0.2       |
| TDG (Gal-β1-4Glc) | 0.02               |
| LacNAc (Gal-β1-4GlcNAc) | 0.02–0.05     |
| Gal-β1-3GlcNAc | 0.02–0.05         |
| Gal-β1-3GlcNAc | 0.05               |
| Gal-β1-6GlcNAc | 0.02–0.05          |
| β-p-AP-Lac<sup>a</sup> | 0.05         |
| Trisaccharides |                  |
| 2-Fuc-Lac    | 0.5–1.0             |
| 3-Fuc-Lac    | 1.0                 |
| 6-Fuc-Lac    | 1.0                 |

<sup>a</sup> AP indicates aminophenyl.
teins in the lactose extracts prepared from other annelids, i.e. earthworm (Oligochaeta), lugworm (Polychaeta), and leech (Hirudinea).

During the course of this study, Cole and Zipser (24, 25) reported the presence of similar galactose-specific lectins, LL16, LL35, and LL63, from the leech Hemopis marmorata. LL16 and LL35 seem to have biochemical properties similar to those of the earthworm lectin investigated in the present study, i.e. in terms of molecular weight, metal independence for sugar binding, and specificity for simple saccharides. As they used detergent to extract leech lectins, it should have solubilized endogenous ligand glycoconjugates to which the lectins were bound. They also reported that LL16 and LL35 showed broader sugar binding specificity than LL63; binding of the former

![Figure 5](image-url)

**FIG. 5.** Nucleotide and deduced amino acid sequences of EW29, clone 5. Solid lines with arrowheads represent the locations of gene-specific primers used for nucleotide sequencing (−1F, 247F, 360R, 405R, and 793R), and dotted lines with arrowheads represent degenerate primers used for the first genomic PCR (Lys-4F, Lys-7R). Potential poly(A) addition signal “AATAAA” is wavy underlined. Dots above nucleotides denote substitution positions observed between the analyzed cDNA clones. Three amino acids, which are the result of “non-silent” substitutions, are shaded (see Table V for details).

| Primer | Length (variation) | Sequence |
|--------|-------------------|----------|
| Lys-4F | 20 (64)           | CA(A/G) (G/A(T/C(GC(T/C(A/G)CA(A/G)CA(A/G)TGGAA)  |
| Lys 4R | 20 (64)           | TTCCA(T/C(TGC(T/G(A/G(T/C(GC(A/G(T/C(T/G(T/G)AT  |
| Lys-7F | 20 (64)           | AA(T/C(CA(A/G)CA(A/G)TGGGA(A/GTC(T/G(A/G(T/G(T/G)GT  |
| Lys-7R | 20 (64)           | ACAG(A/G(T/C(TGC(TCC)CCTC(T/G(T/G(T/G(T/G)AT  |
| Lys-8F | 20 (96)           | ATGCAT(T/C(TGC(A/G)AA(A/G)ATC(T/C(A/G)TGGT(G/A(T/C(A/G)CA  |
| Lys-8R | 20 (96)           | TG(A/G(T/C(TGC(A/G(T/TTT(T/C(TGC(A/G(T/CAT  |
| Lys-10F| 20 (32)           | AA(T/C)CA(A/G)GCC(TG(T/C(A/G)TT(T/C(GA  |
| Lys 10R| 20 (32)           | TC(G/A)AA(A/G(T/C)GTCCACGGC(T/C(T/G(A/G)TT |

| Primer | Length (variation) | Sequence |
|--------|-------------------|----------|
| −1F (Wh-5’) | 22 | GATGCTGGAAGGCTTFTCTTG |
| 247F   | 24 | GTCCGCAAGAGTCTGTCGACCA |
| 360R   | 24 | GTGTCCCTGTGTTATCCAAGG |
| 384F (Wh-5’) | 25 | CATGAGGCGAGTTCTTCAT |
| 405R (Wh-3’) | 24 | TTAGAAAGCTTCTGCTGAGG |
| 793R (Wh-3’) | 23 | CCAGTGGAGTTTACTGGAAG |

**TABLE IV**

| Oligonucleotide primers used in this study | Primer Length (variation) | Sequence |
|-------------------------------------------|--------------------------|----------|
| 1) Degenerate primers used for PCR cloning | Lys-4F 20 (64)           | CA(A/G) (G/A(T/C(GC(T/C(A/G)CA(A/G)CA(A/G)TGGAA)  |
|                                           | Lys 4R 20 (64)           | TTCCA(T/C(TGC(T/G(A/G(T/C(GC(A/G(T/C(T/G(T/G)AT  |
|                                           | Lys-7F 20 (64)           | AA(T/C(CA(A/G)CA(A/G)TGGGA(A/GTC(T/G(A/G(T/G(T/G)GT  |
|                                           | Lys-7R 20 (64)           | ACAG(A/G(T/C(TGC(TCC)CCTC(T/G(T/G(T/G(T/G)AT  |
|                                           | Lys-8F 20 (96)           | ATGCAT(T/C(TGC(A/G)AA(A/G)ATC(T/C(A/G)TGGT(G/A(T/C(A/G)CA  |
|                                           | Lys-8R 20 (96)           | TG(A/G(T/C(TGC(A/G(T/TTT(T/C(TGC(A/G(T/CAT  |
|                                           | Lys-10F 20 (32)          | AA(T/C)CA(A/G)GCC(TG(T/C(A/G)TT(T/C(GA  |
|                                           | Lys 10R 20 (32)          | TC(G/A)AA(A/G(T/C)GTCCACGGC(T/C(T/G(A/G)TT |

| Primer | Length (variation) | Sequence |
|--------|-------------------|----------|
| Lys-4F | 20 (64)           | CA(A/G) (G/A(T/C(GC(T/C(A/G)CA(A/G)CA(A/G)TGGAA)  |
| Lys 4R | 20 (64)           | TTCCA(T/C(TGC(T/G(A/G(T/C(GC(A/G(T/C(T/G(T/G)AT  |
| Lys-7F | 20 (64)           | AA(T/C(CA(A/G)CA(A/G)TGGGA(A/GTC(T/G(A/G(T/G(T/G)GT  |
| Lys-7R | 20 (64)           | ACAG(A/G(T/C(TGC(TCC)CCTC(T/G(T/G(T/G(T/G)AT  |
| Lys-8F | 20 (96)           | ATGCAT(T/C(TGC(A/G)AA(A/G)ATC(T/C(A/G)TGGT(G/A(T/C(A/G)CA  |
| Lys-8R | 20 (96)           | TG(A/G(T/C(TGC(A/G(T/TTT(T/C(TGC(A/G(T/CAT  |
| Lys-10F| 20 (32)           | AA(T/C)CA(A/G)GCC(TG(T/C(A/G)TT(T/C(GA  |
| Lys 10R| 20 (32)           | TC(G/A)AA(A/G(T/C)GTCCACGGC(T/C(T/G(A/G)TT |

| Primer | Length (variation) | Sequence |
|--------|-------------------|----------|
| Lys-4F | 20 (64)           | CA(A/G) (G/A(T/C(GC(T/C(A/G)CA(A/G)CA(A/G)TGGAA)  |
| Lys 4R | 20 (64)           | TTCCA(T/C(TGC(T/G(A/G(T/C(GC(A/G(T/C(T/G(T/G)AT  |
| Lys-7F | 20 (64)           | AA(T/C(CA(A/G)CA(A/G)TGGGA(A/GTC(T/G(A/G(T/G(T/G)GT  |
| Lys-7R | 20 (64)           | ACAG(A/G(T/C(TGC(TCC)CCTC(T/G(T/G(T/G(T/G)AT  |
| Lys-8F | 20 (96)           | ATGCAT(T/C(TGC(A/G)AA(A/G)ATC(T/C(A/G)TGGT(G/A(T/C(A/G)CA  |
| Lys-8R | 20 (96)           | TG(A/G(T/C(TGC(A/G(T/TTT(T/C(TGC(A/G(T/CAT  |
| Lys-10F| 20 (32)           | AA(T/C)CA(A/G)GCC(TG(T/C(A/G)TT(T/C(GA  |
| Lys 10R| 20 (32)           | TC(G/A)AA(A/G(T/C)GTCCACGGC(T/C(T/G(A/G)TT |

| Primer | Length (variation) | Sequence |
|--------|-------------------|----------|
| Lys-4F | 20 (64)           | CA(A/G) (G/A(T/C(GC(T/C(A/G)CA(A/G)CA(A/G)TGGAA)  |
| Lys 4R | 20 (64)           | TTCCA(T/C(TGC(T/G(A/G(T/C(GC(A/G(T/C(T/G(T/G)AT  |
| Lys-7F | 20 (64)           | AA(T/C(CA(A/G)CA(A/G)TGGGA(A/GTC(T/G(A/G(T/G(T/G)GT  |
| Lys-7R | 20 (64)           | ACAG(A/G(T/C(TGC(TCC)CCTC(T/G(T/G(T/G(T/G)AT  |
| Lys-8F | 20 (96)           | ATGCAT(T/C(TGC(A/G)AA(A/G)ATC(T/C(A/G)TGGT(G/A(T/C(A/G)CA  |
| Lys-8R | 20 (96)           | TG(A/G(T/C(TGC(A/G(T/TTT(T/C(TGC(A/G(T/CAT  |
| Lys-10F| 20 (32)           | AA(T/C)CA(A/G)GCC(TG(T/C(A/G)TT(T/C(GA  |
| Lys 10R| 20 (32)           | TC(G/A)AA(A/G(T/C)GTCCACGGC(T/C(T/G(A/G)TT |

**Earthworm 29-kDa Lectin**

Numbers in parentheses are nucleotide numbers. Underlines denote introduced mutations to make initiation codon ATG in 384F (Ch-5’) and termination codon TTA (antisense) for 405R (Nh-3’).
Earthworm 29-kDa Lectin

TABLE V
Summary of nucleotide substitutions found in cDNA and genomic clones for earthworm 29-kDa lectin
Substituted bases relative to clone 5 are bold for emphasis. Underlined codons correspond to underlined amino acids.

| cDNA | Genome |
|------|--------|
| 1     |        |
| 3,5,10 | 14 Gly | 37 Leu | 22 Glu | 24 Gly | 40 Leu | 42 Tyr | 44 Ala | 51 Gly | 56 Leu | 66 Leu | 69 Ala | 84 Gly | 87 Val | 96 Lys | 100 Leu |
| 7     | 5,14   | GGT TGT TCC GGG CTG TAT GCA AGA CAA CTC ATA TTG GCG GGC GTC AAA CTG |
| 8     | 2,16,19| GGT TGT TCA GGA CTG TAC GCA AGA CAA CTC ATA CTG GCA GGG GGC GGT AAG CTG |
| 9     | 3,4,9  | GGT TGT TCA GGA CTG TAC GCA CTA ATA CTG GCA GGG GGC GGT AAG CTG |
| Consensus | GGT CTG TCA GGA CTG TAC GCA AGA CAA CTC ATA CTG GCA GGG GGC GTC AAG CTG |
| cDNA | Genome |
| 1     | 115* Lys | 120 Thr | 149 Gly | 171 Asn | 178 Gln | 191 Ile | 196 Gln | 197 His | 213 Ile | 225 Val | 228 Arg | 238 Ala |
| 7     | 5,14   | AAA ACC GCC AAC CAG ATC GAA CAA ATC CGA GTA GCA GAA CAA ATC CGA GTA GCC |
| 8     | 2,16,19| AGG ACC GTG AAC CAA ATT GAA CAT ATG CAA GAA CAT ATC CGA GTC GCC |
| 9     | 3,4,9  | AGG ACC GTC AAT CAG ATG GAG CAT ATG CAA GAA CAT ATC CGA GTC GCC |
| Consensus | AAA ACC GCC AAC CAG ATC GAA CAA ATC CGA GTA GCA GAA CAA ATC CGA GTC GCC |

* Numbers indicate position.

Fig. 6. Construction and production of recombinant whole Earthworm 29, A, the recombinant earthworm lectin was expressed as a fusion product with β-galactosidase α-peptide in E. coli TOP10F’ cells as described under “Experimental Procedures.” B, E. coli cells were lysed by sonication, and the obtained extract was applied to an asialofetuin-agarose column (bed volume, 10 ml) equilibrated with EDTA-MEPBS. After extensive washing of the column, the bound protein was eluted with EDTA-MEPBS containing 20 mM lactose. C, SDS-PAGE analysis of both flow-through, retarded, and bound fractions were separated by SDS-PAGE. A, the binding of Nh; there was no adsorption to the column, but the elution was strongly retarded. B, the binding of Ch; the adsorbed protein was eluted with 20 mM lactose from fraction 16.

Fig. 7. Functional analysis of individual N-terminal and C-terminal domains of Earthworm 29. On the basis of recombinant EW29 construct (Fig. 6), both individual domains (Nh and Ch) were produced. E. coli extracts were applied to an asialofetuin-agarose column, and both flow-through, retarded, and bound fractions were analyzed by SDS-PAGE. A, the binding of Nh; there was no adsorption to the column, but the elution was strongly retarded. B, the binding of Ch; the adsorbed protein was eluted with 20 mM lactose from fraction 16.

lectins to asialofetuin-agarose was inhibited by galactosamine and α-methylgalactoside as in the case of our preparation, whereas that of the latter was not inhibited by these saccharides. Since the reported molecular size (35 kDa) of LL35 was close to those of our annelid lectins (29–31 kDa), and specific antisera raised against highly purified EW29 reacted significantly with LL35,3 we consider LL35 to be a leech homologue of the earthworm lectin. Galactose-binding lectins recently identified in N. japonica and Marphysa sanguinea (both belonging to Polychaeta; see Ref. 26) as well as Urechisunicinctus (belonging to Echiuridae; see Ref. 27) may also represent counterpart molecules in these annelids.

Although these annelid lectins were prepared essentially by the same strategy as used for galectins, partial peptide analyses suggested that they are not related to galectins but form another protein family having the key consensus motif Gly-X-X-Gln-X-Trp. In fact, the earthworm lectin could be extracted with both galactose and melibiose as well as lactose, whereas galectins in general have only weak affinity to the former saccharides. Although we added β-mercaptopethanol to the buffer for lectin extraction, it proved to be unimportant, because its removal did not significantly reduce the recovery or activity of the earthworm lectin.

The result of cDNA cloning of EW29 was striking in several aspects. First of all, the lectin proved not to be a member of the galectin family, contrary to our expectation. Moreover, the annelid lectins were suggested to be members of a larger family that includes many other carbohydrate-recognition proteins. At least two groups of proteins have been shown to contain the multiple short consensus motif Gly-X-X-Gln-X-Trp (Fig. 8). The first group consists of hemagglutinins (29–35 kDa) including EW29 investigated in this work and HA33 (a 33-kDa major component of hemagglutinin from C. balticum) type C; see Ref. 28) as well as the B-chain of a series of plant toxins. A number of plant toxins consisting of A-chain (ribosome-inactivating protein) and B-chain (lectin) have been reported, e.g. ricin (or RCA60; see Ref. 29), agglutinin (or RCA120; see Ref. 30) from Ricinus communis and their related homologues in elderberry...
**FIG. 8.** Putative superfamily consisting of carbohydrate-recognition proteins having Gly-X-X Gly-X-Trp motif. A, domain structures of proteins having the consensus motif. The members are classified into Lectin group and Enzyme group. For details, see text. Two consecutive consensus motifs are expressed by thick vertical lines in each unit segment, which consists of 40–50 amino acids. Thin vertical lines in the structures signify assumed boundary regions between the domains. B, alignment of amino acid sequences of the members having the consensus motif. Gly (G), Gln (Q), and Trp (W) contained in the consensus motif are emphasized by white letters on black background. Relatively conserved amino acids within the sequences are shaded. One or two consensus amino acids at each of the several conservative positions are depicted above the sequences. Several amino acids, which form the two galactose-binding domains of ricin (40), are shown in inverted boldface; i.e., Asp22, Gln35, Trp37, Lys40, and Asn46 in the N-terminal CRD, and Asp234, Asn255, and Tyr248, in the C-terminal CRD, respectively.

(Sambucus sieboldina), S. sieboldina agglutinin (31), and sieboldin-b (32). Modeccin (33), abrin (34), and viscumin (35) are also this type of toxin found in other plants. The second group consists of various hydrolytic enzymes (see Fig. 8): (i) glycosidases Oerskovia xanthineolytica β-1,3-glucanase (36) and Streptomyces lividans xylanase A (37), and (ii) proteases *Rarobacter faecitabitus* protease I (38) and a subunit of horseshoe crab coagulation factor G (β subunit is a serine protease; see Ref. 39). Notably, all members of the lectin group are known to be specific for galactose/N-acetylgalactosamine except for *S. sieboldina* agglutinin. Since *S. sieboldina* agglutinin binds to sialic acid that is α2,6-linked to galactose (Siaα2,6Gal), it may be a variant having deviated from an ancestral galactose-specific protein. All of these members including EW29 consist of two homologous domains (14–16 kDa), both of which retain carbohydrate binding activity as far as examined. This fact implies that acquisition of multivalency is critical for preservation of efficient lectin function(s). On the other hand, carbohydrate-binding properties are not known for most of the enzyme group. As a sole exception, the CRD of *R. faecitabitus*
protease I has been shown to have mannose binding activity (38). The possibility is excluded that EW29 exists as a larger conjugate via disulfide bridge(s) like ricin and coagulation factor, because Western blotting detected only the 29-kDa species even when β-mercaptoethanol was omitted throughout the experiment.

X-ray crystallographic study has been carried out for ricin complexed with lactose (40). As described, the ricin B-chain has two galactose-binding sites corresponding to the two homologous domains. Amino acid residues forming the binding sites are as follows: Asp$^{22}$, Gln$^{34}$, Trp$^{37}$, Lys$^{40}$, and Asn$^{46}$ for N-terminal CRD, and Asp$^{234}$, Asn$^{255}$, and Tyr$^{248}$ (shown as inverted boldface letters in Fig. 8). These residues are mapped to the first repeat segment of the N-terminal CRD and to the third one of the C-terminal CRD, respectively. Inclusion of tryptophan (or tyrosine) residues at the seventh position of the consensus motif is also intriguing, because such aromatic residues are frequently found in carbohydrate-recognition sites of various sugar-binding proteins, particularly in galactose-specific lectins (41, 42).

Although total amino acid identities between these carbohydrate-recognition proteins were low (<10%), it is noteworthy that these CRDs had striking similarity in their segment architecture; all of the CRDs consisted of three segments each consisting of 40–50 amino acids (5 kDa) and containing 2 consensus motifs in the latter positions. Since the segment size agrees with a moderate “exon” size (43, 44), it is possible to speculate that the original CRD (15 kDa) was the result of several gene duplication events of this ancestral peptide, or “module,” which is presumed to have had primitive carbohydrate binding activity. According to this scenario, one group of the descendants might have evolved as “lectins” (30 kDa) by acquiring divalency by further duplication of the 15-kDa CRD, whereas the other group might have evolved to “enzymes” by ligation with some peptides having catalytic domain(s). Their CRDs could have been useful, e.g., for facilitated binding to target cells covered by carbohydrate chains, as has been suggested for R. faecalisputus protease I (38). X-ray crystallography studies should support this hypothesis. “Rcin superfamily” or “R-type” lectin (where R indicates for ricin and repeat) seem to be appropriate as the name for this novel superfamily.

The present results on cDNA (genome) cloning showed that EW29 is encoded by multiple genes, i.e., genome polymorphism. The analyzed 8 cDNA and 9 genomic clones completely overlapped except for cDNA clone 2 (Table V), which differed from the major cDNA type (clones 1, 3, 5, and 10) at only two positions. Therefore, the actual number of EW29 genes in a single worm would not seem to far exceed four (i.e., the four types represented by cDNA clones 5, 7, 8 and 9; or genomic clones 1, 5, 2, and 3, respectively; see Table V). Difference among the four types of genes (transcripts) spans over the whole coding region at 29 positions, although most of them (26 cases) were silent substitutions. Even for the non-silent substitutions, they were rather homologous ones, and therefore, the generated three polypeptides would not be expected to show very different properties, although their stability or fine sugar-binding specificity might be different. In this regard, three closely similar bands detected by anti-tubifex antibody in the extraction experiment (Fig. 3) might represent these three polypeptides. On the other hand, the fact that half of the obtained cDNA clones were identical to clone 5 suggests that the corresponding CRD gene is more strongly expressed than other genes (corresponding to cDNA clones 7, 8, and 9). Apparently, further functional studies are necessary to evaluate these variant genes.

The physiological function(s) of the investigated annelid lec-
32. Rojo, M. A., Yato, M., Ishii-Minami, N., Minami, E., Kaku, H., Citores L., Girbes, T., and Shibuya, N. (1997) Arch. Biochem. Biophys. 340, 185–194
33. Ready, M., Wilson, K., Piatak, M., and Robertus, J. D. (1984) J. Biol. Chem. 259, 15252–15256
34. Wood, K. A., Lord, J. M., Wawrynczak, E. J., and Piatak, M. (1991) Eur. J. Biochem. 196, 723–732
35. Gabius, H. J., Walzel, H., Yoshi, S. S., Kraup, J., Kojima, S., Gerke, V., Kratzin, H., and Gabius, S. (1992) Anticancer Res. 12, 669–675
36. Shen, S.-H., Chretien, P., Bastien, L., and Sillaty, S. N. (1991) J. Biol. Chem. 266, 1058–1063
37. Shareck F., Roy, C., Yaguchi, M., Morosoli, R., and Kluepfel, D. (1991) Gene (Amst.) 107, 75–82
38. Shimoi, H., Imura, Y., Obata, T., and Tadenuma, M. (1992) J. Biol. Chem. 267, 25189–25195
39. Seki, N., Muta, T., Oda, T., Iwaki, D., Kuma, K., Miyata, T., and Iwanaga, S. (1994) J. Biol. Chem. 269, 1370–1374
40. Montfort, W., Villafranca, J. E., Menzingo, A. F., Ernst, S. R., Katzin, B., Rutemder, E., Xuong, N. H., Hamlin, R., and Robertus, J. D. (1987) J. Biol. Chem. 262, 5398–5403
41. Rini, J. M. (1995) Annu. Rev. Biophys. Biomol. Struct. 24, 551–577
42. Kolatkar, A., and Weis, W. (1996) J. Biol. Chem. 271, 6679–6685
43. Go, M. (1981) Nature 291, 90–92
44. Blake, C. (1983) Nature 306, 535–537
45. Tai, M.-H., Bheuben, M. B., Autio, D. M., and Zipser, B. (1996) J. Comp. Neurobiol. 371, 235–248
46. Arata, Y., Akimoto, Y., Hirabayashi, J., Kasai, K., and Hirano, H. (1996) Histochem. J. 28, 201–207
47. Chiu, M. L., Jones, J. C. R., and O’Keefe, E. J. (1992) J. Cell Biol. 119, 1689–1700