Structural and biochemical characterization of a new type of lectin isolated from carp eggs

Monica GALLIANO*, Lorenzo MINCHIOTTI*, Monica CAMPAGNOLI*, Alberto SALA*, Livia VISAI*, Angela AMORESANO†, Piero PUCCI†, Annarita CASBARRA†, Marco CAUCI‡, Massimiliano PERDUCA‡ and Hugo L. MONACO‡[1]

*Department of Biochemistry 'A. Castellani', University of Pavia, via Taramelli 3b, 27100 Pavia, Italy, †Department of Organic Chemistry and Biochemistry, University of Naples 'Federico II', via Cynthia 6, 80126 Naples, Italy, and ‡Biocrystallography Laboratory, Department of Science and Technology, University of Verona, Ca Vignal 1, strada Le Grazie 15, 37134 Verona, Italy

A previously unidentified glycoprotein present in the eggs of the carp (*Cyprinus carpio*) was isolated and structurally characterized. The protein binds to a Sepharose 4B matrix and can be eluted with 0.4 M *N*-acetylglucosamine. The protein has an apparent molecular mass of 26686.3 Da. On the basis of gel-filtration chromatography, the protein appears to be present in solution as a monomer. The sequence of its 238 amino acids, the position of its four disulphide bridges and the composition of its single N-linked carbohydrate chain were determined. The lectin shows a very low agglutinating activity for human A-type erythrocytes and interacts with both Gram-positive and -negative bacteria. These latter interactions are inhibited by *N*-acetylglucosamine. A database search shows that its amino acid sequence is similar to that of the members of an invertebrate lectin family that includes tachylectin-1. Tachylectin-1 is present in the amoebocytes of the horseshoe crab, *Tachypleus tridentatus*, and plays a role in the innate defence system of this species. Homologous genes are also present in other fish, having 85 % identity with a gene expressed in the oocytes of the crucian carp (*Carassius auratus gibelio*) and 78 % identity with a gene in the cDNA library of the zebrafish (*Danio rerio*).

Key words: carp, fish-egg lectin (FEL), innate immunity, primary structure, tachylectin-1, tectonin, zebrafish.

INTRODUCTION

Lectins can be defined as proteins that recognize specific carbohydrate structures and thereby agglutinate cells by binding to cell-surface glycoproteins and glycoconjugates [1]. In recent years, the role of the protein–carbohydrate recognition phenomenon has steadily attracted attention, as its importance in a variety of fundamental biological processes became more evident. Multivalent protein–carbohydrate interactions have thus been proposed as the new paradigm for supramolecular assembly and signal transduction [2]. Lectins are found in all types of living organisms, either in soluble or in membrane-bound form. They are usually structurally complex molecules with one or more carbohydrate-recognition domains [3]. Animal lectins can play a variety of physiological roles. In particular, these proteins are crucial in the innate immune system where they bind to the carbohydrates present on the surface of potential pathogens [4]. The innate immune system is present in invertebrates and in vertebrates, but in the former it is the only means of defence against invading pathogens [5]. The lectins that participate in the innate immunity of the horseshoe crab, *Tachypleus tridentatus*, are well characterized at both the structural and functional level [6–13]. A number of fish lectins have also been extensively studied. These include proteins that actively participate in the innate immune system [14–16]. However, such proteins have not yet been characterized in fish eggs. The lectins so far described from this source have the interesting property of binding rather specifically to rhamnose [17,18], a property that is shared by very few species outside the fish family [19]. In the present study, we describe the isolation, structural characterization and binding properties of a novel glycoprotein purified from the eggs of the common carp, *Cyprinus carpio*. The protein binds to bacterial cells. This binding can be inhibited most effectively by *N*-acetylglucosamine and to a lesser extent by other sugars including L-rhamnose. We have determined its complete amino acid sequence, carbohydrate composition and the position of its four disulphide bridges. The polypeptide chain is 238 amino acids long. The first 200 residues show significant sequence similarity with a lectin described in the horseshoe crab, tachylectin-1 that has a function in the innate immune system of the arthropod. Genes coding for homologous proteins are also present in other bony fish.

EXPERIMENTAL

Protein purification

The carp eggs were obtained from adult females and were kept frozen until use. In each preparation, approx. 500 g was thawed overnight in 2 litres of 20 mM Tris/HCl, 50 mM KCl, 5 mM CaCl₂ (pH 7.4) and 0.02 % NaN₃, containing PMSF and was homogenized in a blender for 3 min at 4 °C. After centrifugation at 10 000 **g** for 1 h at 4 °C, the supernatant was filtered through a 10 μm mesh nylon net and applied on to a Sepharose 4B column (2.4 cm × 15 cm) pre-equilibrated with the same buffer. The column was washed until the absorbance monitored at 280 nm was negligible. The bound protein was then eluted with 2 M urea or with 0.4 M D-GlcNAc in the initial Tris buffer, dialysed exhaustively, concentrated and submitted to gel-filtration chromatography using a Superdex 75 column HR 10/30 (Amersham Biosciences, Freiburg, Germany).

Analytical gel filtration

Analytical gel filtration was performed on a TSK G3000 SW column (7.5 mm × 60 cm; Tosohaas, Montgomeryville, PA, U.S.A.) connected to an AKTA Purifier System (Amersham Biosciences) and eluted with 25 mM Tris/HCl (pH 7.8) and 0.2 M NaCl, containing 0.4 M D-GlcNAc at a flow rate of 0.3 ml/min. BSA (67 000 Da), ovalbumin (45 000 Da), chymotrypsinogen A (25 000 Da) and RNase A (13 700 Da) were used as reference proteins.

Amino acid sequence analysis of the native FEL (fish-egg lectin)

Automated Edman degradation and sequence analysis of the native purified FEL (1.0 nmol) was performed on a Hewlett-Packard model G 1000A sequencer connected on-line to a phenylthiohydantoin analyser from the same manufacturer.

CNBr cleavage and peptide separation

The FEL (6 mg) was reduced, denatured and treated with iodoacetic acid [20]. Peptide bond cleavage at the methionine residues was performed by incubating the freeze-dried protein (4.5 mg) with CNBr (Fluka, Buchs, Switzerland), and the peptides were separated by reversed-phase HPLC on a Vydac C_{18} column (4.6 mm × 250 mm, 5 μm; The Separations Group, Hesperia, CA, U.S.A.) connected to an AKTA purifier system. The freeze-dried digests were dissolved in 50 μl of 70 % formic acid + 150 μl of 0.1 % aqueous TFA (trifluoroacetic acid; solvent A), and 100 μl (corresponding to approx. 10 nmol) was applied on to the column and eluted with a linear gradient of 0–70 % acetonitrile/2-propanol (2:1, v/v) in 0.1 % TFA (solvent B) for 140 min at a flow rate of 1 ml/min. Amino acid sequence analysis of the six purified CNBr peptides obtained from the carboxymethylated FEL was performed as described for the intact protein.

Tryptic digestion and peptide separation

Trypsin (Roche) was added to the S-carboxymethylated protein (1.5 mg) or to its purified large CNBr fragments (CB1, CB3, CB5 and CB6, approx. 100 μg) dissolved in 500 μl of 0.1 M NH_4HCO_3 (pH 8.0), to give a final enzyme/substrate ratio of 1:20. Digestion was performed at 37 °C for 3 h and the reaction was stopped by adjusting the pH of the solution to 2.0 with TFA. The peptide mixtures obtained were injected on to the Vydac C_{18} column, equilibrated with 0.1 % aqueous TFA (solvent A), and were eluted with a two-step linear gradient of acetonitrile in 0.1 % aqueous TFA (solvent B): from 0 to 5 % for 2 min and from 5 to 55 % for the second 100 min at a flow rate of 1 ml/min. The separated peptides were vacuum-dried and sequenced as described above. These peptides are given the prefix T followed by a number indicating the position of the peptide from the N-terminus of the intact protein.

Monosaccharide analyses by GC–MS

Purified protein (1 nmol) was incubated in 500 μl of methanolic 1 M HCl at 80 °C for 16 h. Re-N-acetylation of the monosaccharide mixture was performed by adding 500 μl of methanol, 10 μl of pyridine and 10 μl of acetic anhydride and incubating at room temperature (25 °C) for 15 min. Sugars were trimethylsilylated in 200 μl of N,O-bis(trimethylsilyl) acetamide at 70 °C for 15 min. The samples were dried under nitrogen, dissolved in 50 μl of hexane and the supernatant was used for the analyses.

MS

MALDI (matrix-assisted laser-desorption ionization) spectra were recorded on a Voyager DE-PRO MALDI–TOF (time-of-flight) mass spectrometer operating in reflection mode. A mixture of the analyte solution and the appropriate matrix (sinapinic, α-cyano-4-hydroxycinnamic or 2,5-dihydroxybenzoic acid) was applied to the sample plate and dried. The data were analysed using the software provided by the manufacturer and are reported as monoisotopic masses. Electrospray–MS (ES–MS) analyses were performed using a ZQ single quadrupole mass spectrometer. In a typical experiment, 25 μl of an individual HPLC peak was injected directly into the ion source and the data were recorded and processed using standard software. GC–MS analyses were performed on a 5390 MSD quadrupole mass spectrometer equipped with a Hewlett-Packard gas chromatograph. Electron ionization mass spectra were recorded by continuous quadrupole scanning at 70 eV ionization energy. N-linked glycans were released from the peptide mixtures when incubated overnight at 37 °C with 0.1 enzyme units of peptide N-glycosidase F in 50 mM ammonium bicarbonate (pH 8.5). The glycan moiety was separated from the peptides by reversed-phase chromatography on pre-packed Sep-Pak cartridges equilibrated with 5 % (v/v) acetic acid. The oligosaccharides were dissolved in water for MALDI-MS analyses.

Two-dimensional and SDS/PAGE analysis

Two-dimensional polyacrylamide gel electrophoresis was performed using the immobilized pH gradient system [21]. The first dimension, isoelectric focusing, was performed on laboratory-made gels, cast on Gel-Bond with a 4–10 non-linear immobilized pH gradient obtained with Acrylamido buffer solutions (Fluka) and the separation was run in the Multiphor II horizontal system (Amersham Biosciences). The carp tissue homogenates obtained from oocytes, liver, kidney, heart, testis and serum were diluted 1:1 with a solution containing 8 M urea, 4 % (w/v) CHAPS, 40 mM Tris, 65 mM dithioerythritol and 0.5 % (v/v) 3.5–10 Ampholine. After centrifugation, 30 μl of each sample was loaded. The gel strips were then equilibrated with SDS, placed on top of vertical 17 % gels, and the second dimension was performed using a Mini PROTEAN II cell (Bio-Rad Laboratories, Hercules, CA, U.S.A.). SDS/PAGE was performed by the method of Laemmli [22] and the gels were stained with Coomassie Blue or silver nitrate.

Antibody production and Western-blot analysis

The purified FEL was emulsified in aluminium hydroxide adjuvant, and 100 μg aliquots were injected intraperitoneally in Balb/c mice five times at 1 week intervals. The mice were bled, and the sera were tested for reactivity to the purified FEL using ELISA and Western-blot analysis. The specific IgGs were purified from plasma by chromatography on a Protein-A–Sepharose column (Amersham Biosciences) according to the manufacturer's recommendations. Antibody titres were assayed either by ELISA or immunoblotting. For Western-blot analyses, electrophoreses were run on SDS/PAGE (17 % gel) and the proteins were transferred by electroblotting to Immobilon-P PVDF membranes (Millipore Corporation, Billerica, MD, U.S.A.) using a Mini Protean II apparatus (Bio-Rad). The membrane was blocked with

5 % (w/v) skimmed milk and probed with the mouse antiserum diluted 1:1000. Immunoreactive spots were detected with horse-radish peroxidase-conjugated anti-mouse immunoglobulin and developed by the enhanced chemiluminescence method (ECL® system; Amersham Biosciences).

Haemagglutinating activity

Trypsinized and glutaraldehyde-treated types B and O and glutaraldehyde-treated type A human erythrocytes were pur-chased from Sigma as 4 % suspensions in PBS. The latter were trypsinized before use.

Binding to bacterial cells by ELISA

Bacteria (*Escherichia coli* strain 89155, *Proteus mirabilis* and *Staphylococcus aureus* Cowan 1 Δ*spa*::TcR) were grown in liquid medium (Brain Heart Infusion; Difco, Detroit, MI, U.S.A.) for 16 h. The cells were collected by centrifugation and suspended in PBS [10 mM sodium phosphate (pH 7.4) containing 0.13 M NaCl] to a final density of 1×10^{10} cells/ml. Cell densities were estimated by comparing the absorbance of the sample with a previously prepared standard curve relating D_{600} to the cell number determined by counting cells in a Petroff Hausser chamber. Microtitre wells were coated overnight at 37 °C with 50 μl of a 1×10^9/ml bacterial suspension of each strain, previously inactivated at 88 °C for 5 min, in 50 mM sodium carbonate (pH 9.5). To block additional protein-binding sites, the wells were treated for 1 h at 22 °C with 200 μl of 2 % (w/v) BSA in PBS and then washed five times with PBST [PBS with 0.1 % (v/v) Tween 20]. For the FEL-binding assay, 100 μl of a 10 μg/ml lectin solution in PBS was added to the coated wells and incubated for 90 min at 22 °C. Plates were then extensively washed with PBST (phosphate-buffered saline Tween 20) and incubated for 90 min with 100 μl of a 10 μg/ml solution of mouse anti-FEL IgG dissolved in PBS with 2 % BSA. After washing five times with PBST, the microtitre wells were incubated for 1 h with rabbit anti-mouse IgG conjugated to horseradish peroxidase (1:1000 dilution; Dako, Gostrup, Denmark). Finally, the conjugated enzyme was reacted with the *o*-phenylethylenediamine dihydrochloride (Sigma) and the absorbance at 492 nm was monitored with a micro-plate reader (Bio-Rad).

Bacterial agglutination and antimicrobial activity

E. coli strain 89155 and *S. aureus* Cowan 1 Δ*spa*::TcR, grown as described above, were used for antimicrobial and bacterial agglutinating activities, which were determined as described by Saito et al. [9].

Competition assays

For the inhibition assays, 100 μl of lipopolysaccharide (*E. coli* 0111), fetuin, asialofetuin, heparin and decorin, at a 100 μg/ml concentration or 250 mM in the sugars, L-fucose, D-galactose, D-glucose, D-mannose, L-rhamnose, D-GlcNAc, D-GalNAc and *N*-acetyllactosamine (Sigma) were preincubated with 100 μl of a 10 μg/ml lectin solution in PBS for 30 min at 22 °C. The mixture was then added to the wells coated with the bacterial cells, incubated further for 90 min at 22 °C and the rest of the ELISA procedure was as described in the previous section. The concentration that gave 50 % inhibition (K_i) was calculated for D-GlcNAc using the residual binding to bacteria measured
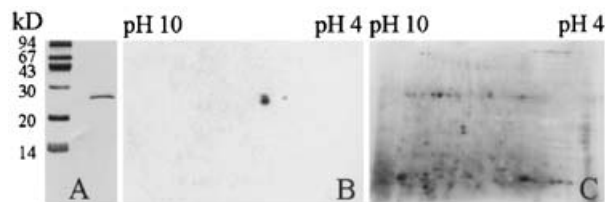

Figure 1   Electrophoretic analyses

(**A**) Purified protein subjected to SDS/PAGE (17 % gel). The gel was stained with Coomassie Blue and the masses of the standard proteins are indicated on the left. (**B**) Two-dimensional electrophoresis of the total carp egg homogenate transferred by electroblotting and developed with antiserum against the FEL. (**C**) Silver-stained two-dimensional electrophoresis of the total egg homogenate.

after preincubating the lectin with different sugar concentrations in the range 0–250 mM.

Determination of affinity constant for D-GlcNAc–BSA

Microtitre plates were coated with 100 μl of GlcNAc coupled with BSA (Sigma) at 30 μg/ml concentration in 50 mM sodium carbonate (pH 9.5). The plates were then blocked with BSA for 1 h at 22 °C and washed extensively. Increasing concentrations of FEL were then added to the coated wells that were incubated further for 90 min at 22 °C. The rest of the ELISA procedure was as described in the bacteria-binding assay. The final absorbance, read at 492 nm, was used in the following equation to calculate the K_a value:

$$A = A_{max}[FEL]K_a/(1 + K_a[FEL])$$

RESULTS AND DISCUSSION

Purification and analytical gel filtration of the carp egg lectin

The lectin was purified essentially in a single step by using its unusual property of binding selectively to Sepharose 4B. The protein was initially eluted with the buffer containing 2 M urea and, after its lectin nature was ascertained, with 0.4 M D-GlcNAc. The use of a bacteriostatic agent (0.02 % sodium azide) ensured yields that were approx. 10 mg of pure protein for 500 g of fish eggs. The purified lectin showed one band with apparent molecular mass of 27000 Da in SDS/PAGE (see Figure 1A) and a single N-terminal amino acid, leucine, when submitted to Edman degradation. A Western-blot analysis of the total egg homogenate probed with the anti-FEL antiserum shows one predominant spot with pI 4.9 migrating as three closely spaced bands in the second direction and a minor isoform at pI 4.5 (Figure 1B). Figure 1(C) shows the silver-stained two-dimensional electrophoresis of the same homogenate for comparison. No positive spots were detected in the other carp tissues examined or in the serum in Western-blot analysis.

On gel filtration, using a Superdex 75 column, it eluted as a single sharp peak with a retention volume indicative of the monomer mass. To exclude the possibility that the lectin might have been retarded in the polysaccharidic matrix of the Superdex 75 resin, two other experiments were performed. The first was to include 0.4 M D-GlcNAc in the Superdex 75 elution buffer. The result was that the elution volume did not change significantly. The second was to use a different type of gel. With the silica-based TSK G3000 SW resin, the apparent molecular mass, calculated
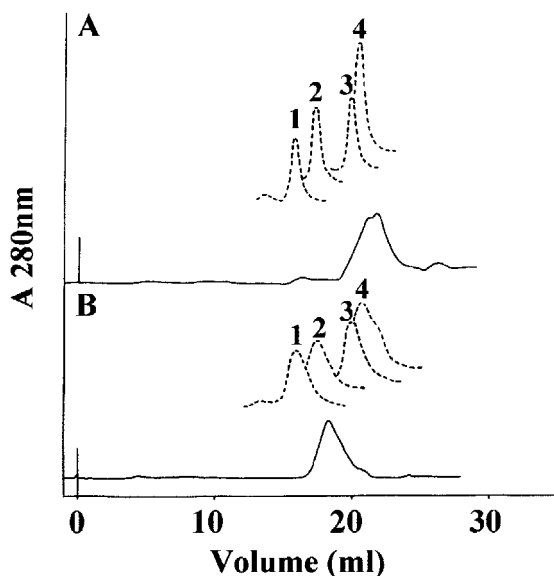
Figure 2   Analytical gel filtration

Apparent molecular mass of the FEL on gel filtration on a TSK G3000 SW column eluted with 25 mM Tris/HCl (pH 7.8) and 0.2 M NaCl (**A**), containing 0.4 M D-GlcNAc (**B**). The reference proteins (- - - -) are, in Da: (1) BSA (67 000), (2) ovalbumin (43 000), (3) chymotrypsinogen A (25 000) and (4) RNase A (13 700).

Figure 3   Summary of the complete amino acid sequence determination of the carp egg lectin

The sequenced regions are marked by arrows. N-t represents the N-terminal sequence of the native protein, and the origins of the peptides are designated by CB for CNBr and T for tryptic hydrolysis.

on the basis of the retention volume, was approximately half the value predicted for a single polypeptide chain. This indicated that the lectin was bound to the resin and that the estimated value was unreliable. Therefore 0.4 M D-GlcNAc was added to the elution buffer to prevent non-specific binding, and FEL eluted with a volume corresponding to the monomer mass slightly increased by the presence of the bound carbohydrate (Figure 2).

Amino acid sequence

Positions of the peptides and the overall strategy used for the complete amino acid sequence determination of the FEL are summarized in Figure 3. The N-terminal sequence analysis of the native protein established the initial 67 N-terminal residues. CNBr digestion of the reduced and carboxymethylated FEL produced six peptides (CB1–CB6), which were isolated and submitted to automated Edman degradation. CB2 and CB4 were sequenced through their C-termini (they were 9 and 23 residues long respectively). The complete primary structure of the four larger CB fragments was obtained by sequence analysis of their tryptic peptides, isolated by reversed-phase HPLC and sequenced through their C-termini. Alignment of the six CB fragments was achieved using the overlapping tryptic peptides. Tryptic digestion of S-carboxymethylated FEL originated 14 peptides (T1–T13 and T15) plus free lysine (T14), which were isolated by reversed-phase HPLC. Peptides T6, T9 and T12 were sequenced through their C-termini: T6 was 29-amino-acids long and over-lapped with CB1, CB2 and CB3; T9 was 22-amino-acids long and overlapped with CB3 and CB4; and T12 was 19-amino-acids long and overlapped with CB5 and CB6. The fourth overlapping peptide, T10, was obtained in low yield, probably because of its very low solubility but its initial 25 N-terminal residues overlapped with CB4 and CB5. The CNBr and tryptic peptide mixtures were also analysed by MALDI-MS to check the sequences that had been established chemically and the primary structure of the protein was confirmed. This analysis also
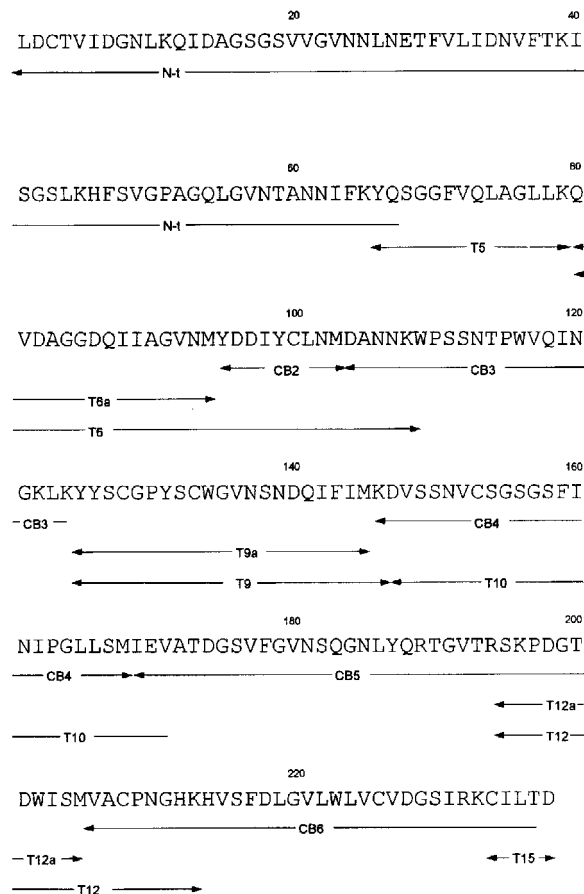
showed that the molecular mass of T2 as well as that of CB1 (1–94) was higher by 1216.3 Da when compared with the value predicted by the amino acid sequence. The existence of one canonical Asn-X-Thr tripeptide acceptor sequence for N-glycosylation in position 27 suggested that the explanation was the presence of a glycosidic moiety in this part of the sequence.

Carbohydrate composition

Owing to the presence of heterogeneous glycans attached to the polypeptide chain, the ES mass spectrum of the FEL showed a series of peaks that can be assigned to different molecular mass forms of the protein. The different glycoforms were identified on the basis of the theoretical mass of the polypeptide chain of the protein (25 477.6 Da) and the oligosaccharide structures that can be expected from the biosynthetic pathways of N-linked glycans. Table 1 reports the molecular mass, composition and structure type of N-linked glycosidic structures. The highest peak at the molecular mass of 26 686.3 Da corresponds to the protein carrying a $Hex_5HexNAc_2$ glycan structure. The same oligosaccharide structure explained the increase in molecular mass owing to the presence of the glycan in peptides T2 and CB1. The determination of the accurate

Table 1   N-linked glycosidic structures observed in the ES–MS spectra of the native FEL

Molecular mass (Da)	Composition	Structure type
26 686.3	FEL + Hex$_5$HexNAc$_2$	High mannose
26 728.2	FEL + Hex$_4$HexNAc$_3$	Hybrid
27 052.2	FEL + Hex$_6$HexNAc$_3$	Hybrid
27 093.6	FEL + Hex$_5$HexNAc$_4$	Hybrid
27 134.8	FEL + Hex$_4$HexNAc$_5$	Hybrid/complex
27 255.3	FEL + Hex$_6$HexNAc$_4$	Hybrid/complex
27 296.4	FEL + Hex$_5$HexNAc$_5$	Hybrid/complex
27 459.1	FEL + Hex$_6$HexNAc$_5$	Hybrid/complex
27 662.7	FEL + Hex$_5$HexNAc$_6$	Hybrid/complex
27 703.9	FEL + Hex$_6$HexNAc$_6$	Hybrid/complex

Table 2   The disulphide bridge pattern of carp FEL compared with that of tachylectin-1 [9]

Carp FEL	Tachylectin-1
3–234	32–36
100–153	108–112
128–133	183–187
208–226	

molecular mass of the different components (Table 1) showed that all the protein isoforms displayed the glycan chain and therefore it can be concluded that the Asn^{27} site is fully glycosylated. The monosaccharide composition of the glycosidic moiety was determined by GC–MS analyses. The total ion current profile showed a high mannose content and low abundance of galactose and *N*-acetylglucosamine, which is consistent with the presence of high mannose and complex or hybrid type N-linked glycans. No GalNAc residues were observed which rules out the possible presence of O-linked oligosaccharides.

Disulphide pairing

The disulphide bridge assignment was performed combining chemical and enzymic hydrolyses of the native protein with different MS procedures. The native protein (2 nmol) was digested with CNBr and the peptide mixture was directly analysed by MALDI-MS, leading to the identification of the first two disulphide bridges. The MALDI spectrum has three signals related to S–S containing peptides detected at 4775.7, 3592.5 and 3363.4 Da respectively. The fourth and highest peak, at 3937.2 Da, corresponds to the peptide 169–205 which does not contain any cysteine residues. The signal at 4775.7 Da, occurring at 2 Da lower than the expected value for peptide 104–145, supports the existence of an S–S bridge joining the two cysteine residues 128 and 133 in the peptide. The signal at 3363.4 Da was assigned to the peptide pair 95–103 and 146–168 linked by the disulphide bridge Cys^{100}–Cys^{153} and the signal at 3592.5 Da to the peptide 206–238 that contains three cysteine residues. This last signal is found at a mass value 2 Da lower than the theoretical value, again indicating the existence of an internal disulphide bridge. Assignment of the last two bridges required the separation of the peptides in the mixture by HPLC. The isolated peptides were analysed by ES–MS. This analysis showed a peak with molecular mass 14476.5 ± 0.9 Da identified as the peptide pair 1–94 (which contains the sugar chain) linked to the peptide 206–238. These two peptides contain the remaining four cysteine residues in the molecule, i.e. Cys^3, Cys^{208}, Cys^{226} and Cys^{234}. To pair the cysteine residues in these peptides, the fraction that had been isolated by HPLC was subdigested with trypsin and again analysed by MALDI-MS. The enzymic cleavage separated the peptide cluster into two peptide pairs giving rise to signals at 1970.9 and 1752.6 Da. These were assigned to the peptide 206–213, linked to the peptide 223–233 through the S–S bridge Cys^{208}–Cys^{226}, and to the peptide pairs 1–11 + 234–238, linked through Cys^3 and Cys^{234}. These assignments were confirmed by dithiothreitol incubation and manual Edman degradation coupled

with MALDI-MS. The complete pattern of disulphide bridges present in the molecule is presented in Table 2.

Biochemical properties

The lectin-like nature of the purified protein was qualitatively examined by absorbing aliquots (300 μg) to a 1.5 ml Sepharose 4B column and eluting with different monosaccharide gradients (results not shown). In these assays, the protein was eluted effectively by 0.4 M *N*-acetylglucosamine and up to 1.0 M concentrations of glucose, galactose or mannose. Although the presence of 5 mM Ca^{2+} increased the affinity for the resin, EDTA was not effective to elute the protein which indicates that the metal ion is not essential for binding to the Sepharose matrix.

The lectin was tested for haemagglutinating activity with all types of human erythrocytes. It only showed a very low activity for type A erythrocytes, which was neither Ca^{2+}- nor EDTA-dependent. Binding assays were performed with several strains of Gram-positive and -negative bacteria. The lectin has the highest affinity for *E. coli* and *S. aureus* Cowan 1 Δspa::Tc^R, in a dose-dependent and saturable manner, and shows only a negligible interaction with *P. mirabilis* (Figure 4A). The affinity of these interactions was increased by the addition of up to 5 mM Ca^{2+} ions, but it was not abolished by EDTA. The results of growth inhibition experiments against *E. coli* and *S. aureus* are reported in Table 3 and show that the lectin does not have significant antimicrobial activity. No bacterial agglutination was observed after the addition of up to 120 μg/ml of FEL to both strains. These results indicate that the lectin binds with a rather broad specificity to the glyco-structures exposed on Gram-positive and -negative bacteria, but has neither antimicrobial nor agglutinating activity.

Sugar binding specificity

The specificity of sugar binding was investigated using a competition-binding assay with *S. aureus* and *E. coli* in which the protein was preincubated with different carbohydrates at 250 mM concentration. Figure 4(B) shows that, in agreement with the results obtained with the column binding assay, D-GlcNAc was the most potent inhibitor, whereas the non-acetylated saccharide was not effective. D-GlcNAc showed an inhibition of up to 70 and 80% with *S. aureus* and *E. coli* respectively, with a 50% inhibitory concentration of 20 mM (Figure 4C). The interaction with immobilized bacteria was inhibited up to 30 and 40% by D-GalNAc, whereas non-acetylated galactose did not show any effect. Among the non-acetylated monosaccharides, L-rhamnose and L-fucose showed only a weak inhibition of up to 10 and 20% respectively. Binding was not influenced by the presence of lipopolysaccharide or the tested glycoproteins (such as fetuin). These results indicate that (1) the FEL binds preferentially to *N*-acetylglucosamine, (2) it distinguishes between sugars
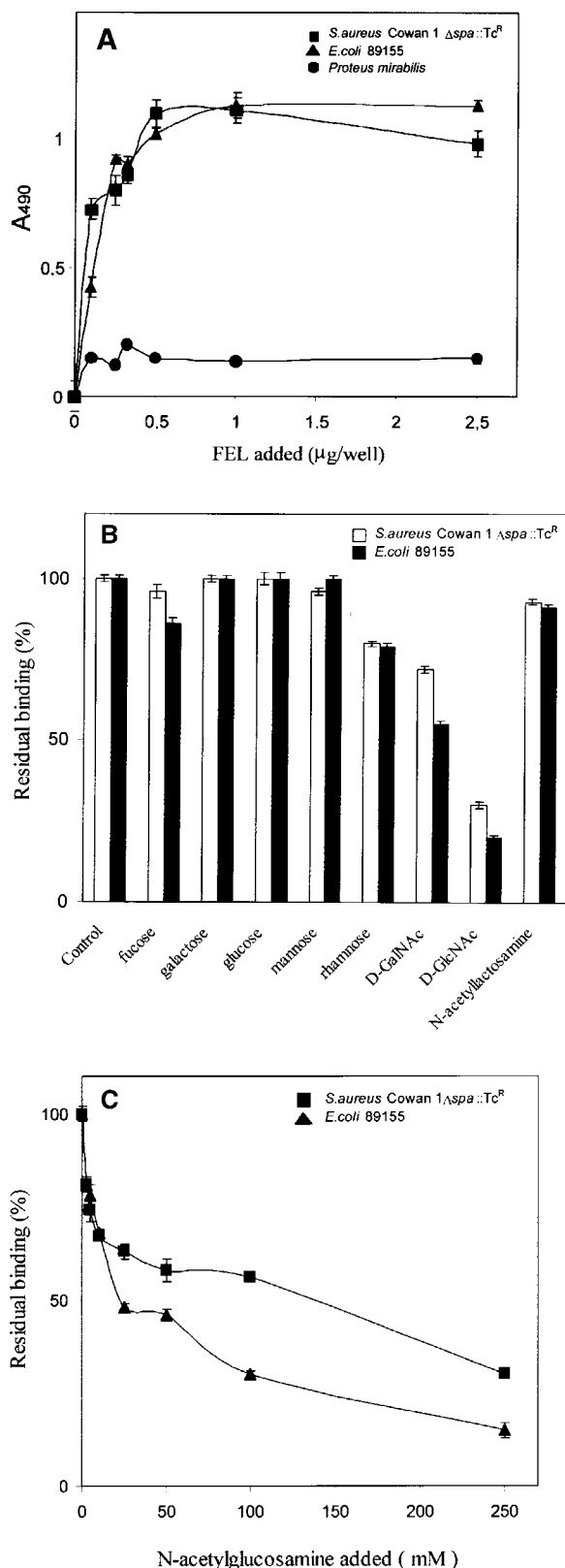
Table 3   Antibacterial activity of FEL, measured in terms of viable cells/ml, against *S. aureus* Cowan 1 △*spa*::Tc^R and *E. coli* 89155

S.D. were calculated from the results of three experiments.

Concentration (μg/ml)	*S. aureus*	*E. coli*
0	$18\,000 \pm 75$ (100%)	$21\,000 \pm 75$ (100%)
40	$14\,000 \pm 50$ (81%)	$19\,000 \pm 65$ (90%)
80	$12\,400 \pm 60$ (69%)	$17\,400 \pm 55$ (81%)
120	$11\,000 \pm 85$ (61%)	$14\,000 \pm 70$ (67%)

containing or not containing the acetamido group and (3) it recognizes the absence of the hydroxy group from C6. For a more direct and quantitative assessment of the affinity of the FEL for its most specific ligand, GlcNAc coupled with BSA was immobilized on the microtitre plates and the specific binding activity of the lectin was measured by the ELISA. The assay yielded an association constant $K_a = 1 \times 10^7$ M for the interaction of the lectin with GlcNAc–BSA.

Sequence homology with members of an invertebrate lectin family

The plasma and haemocytes of the horseshoe crab, *T. tridentatus*, contain several lectins involved in the protection mechanism of the innate immune system of the arthropod. Two of these proteins, tachylectin-1, formerly called lectin L-6 [9], present in the haemocytes and TPL-1, tachypleus plasma lectin 1 [13], are highly homologous since they present a 65% identity in their amino acid sequence. The two proteins are, in turn, homologous with two other lectins, characterized in the plasmodium *Physarum polycephalum*, which are called Tectonins I and II and are located in the exterior surface of the plasma membrane [23]. A characteristic of the amino acid sequences of the members of this protein family is the presence of six internal tandem repeats, which has prompted the suggestion that the proteins are structured as β-propeller domains [24]. Figure 5(A) shows the sequences of these four proteins aligned together with the sequence of the carp egg lectin that we present here. Amino acids that have been boxed are those that are identical in the carp lectin and in at least one of the other four proteins. Note that, with the exception of approximately the last 40 amino acids of the carp lectin, the first part of the sequence, extending for approx. 200 amino acids, is significantly homologous with the sequence of the members of this protein family. The amino acids that are boxed in Figure 5(A) are 95, which is 47.5% of 200. This similarity suggests that the carp FEL is structurally and possibly functionally related to the invertebrate lectin family. This family includes the horseshoe crab proteins that play a role in the innate immune system. In Figure 5(B), it can be better appreciated that there are internal repeats in the carp egg lectin as well. In Figure 5(B), the repeated amino acids are indicated in boldface. Of the eight cysteine residues present in the sequence of the carp FEL, only one, Cys[153], is found in the aligned sequences of tachylectin-1 and TPL-1. The disulphide bridge patterns of FEL and tachylectin-1, presented for comparison in Table 2, are totally different. It is also worth mentioning that the horseshoe crab lectin L-6 (tachylectin-1) is a monomer with affinity for a Sepharose CL-6B matrix, and a rather broad sugar binding specificity [9]. It shows activity against Gram-negative bacteria and is known to participate in the innate immune system of the arthropod. Interestingly, another highly homologous lectin, named TL-P, isolated from the perivitelline fluid of the same species, although

Figure 4   Carp egg lectin binding to bacterial cells

(**A**) Dose-dependent binding of the FEL to microtitre wells coated with bacterial cells. (**B**) Determination of the carbohydrate binding specificity. Equal amounts of different sugars (250 mM) were preincubated with the FEL, and the residual binding to the microtitre wells coated with bacteria was assayed. (**C**) Residual binding to bacteria determined by competitive assay with increasing concentrations of D-GlcNAc.

A

```
carp FEL   LDCTVIDGNL  KQIDA-GSGSV  VGVNNLNETF  VL-IDNV--F  TKISGSLKHF   47
TL1  (L6)  GQWTQISGSL  KQVDA-DDHEV  WGVNRNDDIY  KRPVDGSGSW  VRVSGKLKHV   88
TPL-1      SNWIKVEGSL  KQIDA-DDHEV  WGVNSNDNIY  KRPVDGNGSW  IQIKGGLKHV   87
Tect 1     HKWHKVDGEL  TNISVGHDGEV  WGVNKNHNIY  RLDR-SNNKW  TQIPGELVQV   87
Tect 2     SKWHQIEGAA  TNISVGLDGTV  WCVNKAHEIY  RLDR-GTNKW  SIVPGELVQV  223


Carp FEL   SVGPAGQL-G  VNTANNIFKY  Q---SGGFVQ  LAGLLKQVDA  GGDQ-IIAGVN   93
TL1  (L6)  SASGYGYIWG  VNSNDQIYKC  PKPCNGAWTQ  VNGRLKQID-  GGQS-MVYGVN  137
TPL-1      SASGYGYIWG  VNSKDQIFKC  PKPCNGEWEL  VDGSLKQVD-  GGRD-LVYGVT  136
Tect 1     SVGSHHHVWG  VNHLDHIYKW  DHHHN-KWDK  IDGALTNVS-  VGKDGTVYGVN  136
Tect 2     SVGNSHNIWG  VNHLDAIYKW  NADSN-SWTF  VDGQLTNVS-  VGHDGTVYGVN  272


carp FEL   MYDDIYCLNM  DANNKWPSSN  TPWVQINGKL  KYYS-CGPYS  CWGVNSNDQI  142
TL1  (L6)  SANAIYRRPV  DGSGSWQ---  ----QISGSL  KHITGSGLSE  VFGVNSNDQI  180
TPL-1      QNDEIFRRPV  DGSGVWV---  ----NIPGKL  KHISGSGSWE  VFGVNCNDQI  179
Tect 1     RGHQIYRW--  DGSKVDL---  -----VLGEL  VQIHVSDAEK  IVGVNHLDHI  176
Tect 2     RAGNIYHY--  NGNSWDA---  -----VSGEL  VQIHVANKDL  IVGVNKAGHV  313


carp FEL   FIMKDVSSNVC  SGSGSFINIP  GLLSMIEVAT  DGSVFGVNSQ  GNLYQRTGVT  193
TL1  (L6)  YRC----TKPC  SGQWSLID--  GRLKQCD-AT  GNTIVGVNSV  DNIY-RSG   221
TPL-1      FRC----KKPC  SGQWVRLS--  GYLKQCD-AS  GDSLLGVNSN  DDIFESVPAS  223
Tect 1     YRL------KH  GKDWEKLD--  GELTWVSVGH  HGEVWGVNKL  HHIYKATL   217
Tect 2     YRL------KH  GKDWEKLE--  GELSWVAVGH  GGELWGANSA  HNIYKALL   353


Carp FEL   RSKPDGTDWI  SMVACPNGHK  HVSFDLGVLW  LVCVDGSIRK  CILTD   238
TPL-1      KSCWMNPFL                                              232
```

B

```
carp      LDCTVIDGNL  KQIDAGSGSV  VGVNNLNETF  VLIDNVFTKI  SGSLKHFSVG   50
zebra-f   LDCTIMSGNL  KQIDAGSGSV  VGVNDLNQAF  VLQDDVFNPV  SRSLKHFSVG
sc-c      LNCNVVNGNL  KQIDAGSGSV  VGVNNNNEIF  VLIDNIFTKI  SGSLKHFSVG

carp      PAGQLGVNTA  NNIFKYQSGG  FVQLAGLLKQ  VDAGGDQIIA  GVNMYDDIYC  100
zebra-f   PAGQLGVNKN  NYIFKLMSGR  FVEFPGLLKQ  VDAGGDQIIA  GVNMNDDIFC
sc-c      PAGQLGVNTA  NNIFRFQSGG  FVRLEGLLNQ  VDAGGDQIIA  GVNMYDDIFC

carp      LNMDANNKWP  SSNTPWVQIN  GKLKYYSCGP  YSCWGVNSND  QIFIMKDVSS  150
zebra-f   LNMDASNQWP  SSTTPWVTIN  GKLKYYSCGP  YSCWGVNSDD  YIFMMKGVSS
sc-c      SNMDANNKWL  SSNIPWINIG  GKLKYYSCGP  YSCWGVNHND  QIFIMKDVSS

carp      NVCSGSGSFI  NIPGLLSMIE  VATDGSVFGV  NSQGNLYQRT  GVTRSKPDGT  200
zebra-f   NACSGT--FV  NIPGLLSMIE  VGTDGSVFGV  NYEAKLYQRV  GVSRSNPAGT
sc-c      SVCSGSGSFV  NIPGLLSMIE  VATDGSVYGV  NSQGSLFKRT  GVTRCTPDGT

carp      DWISMVACPN  GHKHVSFDLG  VLWLVCVDGS  IRKCILTD   238
zebra-f   DWISMIACPI  GHKHVSLDLG  VLWVVCVDGS  IRKCTL
sc-c      DWIPVVACPN  GHQHVSFDLG  VLWVVCVDGS  IRKCS
```

Figure 5   Amino acid sequence homologies of the lectin

(**A**) Comparison with the sequences of the members of an invertebrate lectin family. The amino acids that are identical in the carp FEL and at least one of the other four proteins have been boxed. TL1, tachylectin-1 [9]; TPL-1, tachypleus plasma lectin 1 [13]; Tect 1 and 2, tectonins 1 and 2 [23]. (**B**) Comparison with the proteins coded by the homologous genes of the zebrafish and silver crucian carp (sc-c) [25]. The tandem repeats are in boldface and the amino acids that are identical in all the three proteins are boxed. There are a total of 165, which is 69.3 % of 238. All the eight cysteine residues and the tandem repeats present in the molecule are conserved in the three sequences but the glycosylation sequon is present only in the carp FEL.

sharing 218 out of 221 amino acids with tachylectin-1, has a much lower antibacterial activity (and very similar to that of the carp FEL). TL-P is believed to be involved in the development of the embryo and not in the defence against invading pathogens [11].

Homologous genes in other bony fish

A database search revealed that homologous genes are present in other varieties of fish. The carp protein has 85 % identity with a gene expressed in the oocytes of the silver crucian carp (*Carassius auratus gibelio*) [25] and 78 % identity with a gene present in the cDNA library of the zebrafish (*Danio rerio*). Figure 5(B) shows the three sequences aligned. In this case, the amino acids boxed are those that are identical in all the three proteins. Note that the glycosylation sequon, NET, in positions 27–29 of the FEL is not conserved in the other two species. The crucian carp is a triploid bisexual species that can reproduce both sexually and by gynogenesis. In this second case, the sperm of a related species triggers embryogenesis, but does not make any genetic contribution to the egg. Thus the eggs of this species are of two types. The gene coding for the protein homologous with the lectin described in the present study, is expressed differently in these two

types of eggs [25]. This oocyte-specific protein was suggested to be membrane-bound, with a galactose-binding domain, and probably plays a role in the development of the egg. On the basis of its sequence similarity with FEL, we believe that it is unlikely that it is indeed a membrane protein, but the reasons for the different expression in these two types of eggs undoubtedly merit further study.

In conclusion, our results support the proposal of the existence in bony fish of a protein family with similarities to an invertebrate lectin group that includes members with a function in the innate immune system or in embryonic development. In our laboratories, we have coined and used the word 'fishelectin' (FEL; by analogy with tachylectin) to name the protein we described. This term could perhaps be extended to the other members of this lectin family that may be characterized in the future.

REFERENCES

1   Lis, H. and Sharon, N. (1998) Lectins: carbohydrate-specific proteins that mediate cellular recognition. Chem. Rev. **98**, 637–674

2   Sacchettini, J. C., Baum, L. G. and Brewer, C. F. (2001) Multivalent protein–carbohydrate interactions. A new paradigm for supermolecular assembly and signal transduction. Biochemistry **40**, 3009–3015

3   Dodd, R. B. and Drickamer, K. (2001) Lectin-like proteins in model organisms: implications for evolution of carbohydrate-binding activity. Glycobiology **11**, 71R–79R

4   Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A. and Dwek, R. A. (2001) Glycosylation and the immune system. Science **291**, 2370–2376

5   Hoffmann, J. A., Kafatos, F. C., Janeway, Jr, C. A. and Ezekowitz, R. A. B. (1999) Phylogenetic perspectives in innate immunity. Science **284**, 1313–1318

6   Iwanaga, S., Kawabata, S. and Muta, T. (1998) New types of clotting factors and defense molecules found in horseshoe crab hemolymph: their structure and functions. J. Biochem. (Tokyo) **123**, 1–15

7   Kawabata, S. and Iwanaga, S. (1999) Role of lectins in the innate immunity of horseshoe crab. Dev. Comp. Immunol. **23**, 391–400

8   Iwanaga, S. (2002) The molecular basis of innate immunity in the horseshoe crab. Curr. Opin. Immunol. **14**, 87–95

9   Saito, T., Kawabata, S., Hirata, M. and Iwanaga, S. (1995) A novel type of limulus lectin-L6. J. Biol. Chem. **270**, 14493–14499

10 Saito, T., Hatada, M., Iwanaga, S. and Kawabata, S. (1997) A newly identified horseshoe crab lectin with binding specificity to O-antigen of bacterial lipopolysaccharides. J. Biol. Chem. **272**, 30703–30708

11 Nagai, T., Kawabata, S., Shishikura, F. and Sugita, H. (1999) Purification, characterization, and amino acid sequence of an embryonic lectin in perivitelline fluid of the horseshoe crab. J. Biol. Chem. **274**, 37673–37678

12 Chiou, S. T., Chen, Y. W., Chen, S. C., Chao, C. F. and Liu, T. Y. (2000) Isolation and characterization of proteins that bind to galactose, lipopolysaccharide of *Escherichia coli* and protein A of *Staphylococcus aureus* from the hemolymph of *Tachypleus tridentatus*. J. Biol. Chem. **275**, 1630–1634

13 Chen, S. C., Yen, C. H., Yeh, M. S., Huang, C. J. and Liu, T. Y. (2001) Biochemical properties and cDNA cloning of two new lectins from the plasma of *Tachypleus tridentatus*. J. Biol. Chem. **276**, 9631–9639

14 Magor, B. G. and Magor, K. E. (2001) Evolution of effectors and receptors of innate immunity. Dev. Comp. Immunol. **25**, 651–682

15 Ellis, A. E. (2001) Innate host defense mechanisms of fish against viruses and bacteria. Dev. Comp. Immunol. **25**, 827–839

16 Jones, S. R. M. (2001) The occurrence and mechanisms of innate immunity against parasites in fish. Dev. Comp. Immunol. **25**, 841–852

17 Tateno, H., Saneyoshi, A., Ogawa, T., Muramoto, K., Kamiya, H. and Saneyoshi, M. (1998) Isolation and characterization of rhamnose-binding lectins from eggs of steelhead trout (*Oncorhynchus mykiss*) homologous to low density lipoprotein receptor superfamily. J. Biol. Chem. **273**, 19190–19197

18 Hosono, M., Ishikawa, K., Mineki, R., Murayama, K., Numata, C., Ogawa, Y., Takayanagi, Y. and Nitta, K. (1999) Tandem repeat structure of rhamnose-binding lectin from catfish (*Silurus asotus*) eggs. Biochim. Biophys. Acta **1472**, 668–675

19 Kubo, T., Kawasaki, K. and Natori, S. (1993) Transient appearance and localization of a 26-kDa lectin, a novel member of the Periplaneta lectin family, in regenerating cockroach leg. Dev. Biol. **156**, 381–390

20 Gurd, F. R. N. (1972) Carboxymethylation. Methods Enzymol. **25**, 424–438

21 Gianazza, E., Astrua-Testori, S., Giacon, P. and Righetti, P. G. (1985) An improved protocol for 2D maps of serum proteins with immobilized pH gradients in the first dimension. Electrophoresis **6**, 332–339

22 Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**, 680–685

23 Huh, C. G., Aldrich, J., Mottahedeh, J., Kwon, H., Johnson, C. and Marsh, R. (1998) Cloning and characterization of *Physarum polycephalum* tectonins. Homologues of limulus lectin L6. J. Biol. Chem. **273**, 6565–6574

24 Fülöp, V. and Jones, D. T. (1999) B propellers: structural rigidity and functional diversity. Curr. Opin. Struct. Biol. **9**, 715–721

25 Xie, J., Wen, J. J., Chen, B. and Gui, J. F. (2001) Differential gene expression in fully-grown oocytes between gynogenetic and gonochoristic crucian carps. Gene **271**, 109–116