ELSEVIER

Comparative Biochemistry and Physiology Part B 125 (2000) 279–289

CBP

www.elsevier.com/locate/cbpb

A collectin-like protein from tunicates

Sham V. Nair, Sarina Pearce, Peter L. Green, Deepika Mahajan, Rebecca A. Newton, David A. Raftos *

Department of Biological Sciences, Macquarie University, North Ryde, NSW, 2109, Australia

Received 6 August 1999; received in revised form 16 November 1999; accepted 18 November 1999

Abstract

Collectins are a sub-family of C-type lectins from mammals and birds that are characterized by their collagen-like domains. The mammalian collectin, mannose binding lectin, has attracted considerable interest because it can activate complement components via a lectin-mediated complement pathway that is independent of immunoglobulins. In this study, we have identified a calcium-dependent lectin from the invertebrate (tunicate), *Styela plicata*, that bears substantial similarities to mammalian collectins. The tunicate lectin, which was isolated by carbohydrate affinity chromatography, has a reduced apparent molecular mass of 43 kDa. The 43 kDa reduced polypeptide appeared as dimers, trimers and hexamers when analyzed by non-reducing and two-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis, while gel filtration suggested that the native form of the protein was a nonamer. Amino acid sequence and amino acid composition analysis revealed obvious similarities between the tunicate lectin and mammalian collectins, notably the inclusion of a collagenous domain and a short, cysteine bearing N-terminal domain. The identification of a collectin-like protein in an invertebrate such as *S. plicata*, which does not express immunoglobulin, indicates that lectin-mediated complement pathways may predate the origin of antibodies.

Keywords: Collectin; Lectin; Tunicate; Ascidian; Complement; Immunity; Evolution; Mannose-binding lectin

1. Introduction

Collectins are small family of calcium-dependent (C-type) lectins that incorporate collagen-like domains (Holmskov et al., 1994; Hoppe and Reid, 1994; Epstein et al., 1996; Lu, 1997). Five distinct collectins have so far been identified in mammals and birds: mannose binding lectins (MBL), lung surfactant proteins A and D (sp-A and sp-D), conglutinins and serum collectin-43 (CL-43) (Holmskov et al., 1994). Collectin polypeptides range in apparent molecular mass from 24 to 48 kDa, and are characterized by the expression of four discrete functional domains (Lee et al., 1991; Holmskov et al., 1994). A short (7–25 amino acid) amino terminal 'tail' domain precedes a

Abbreviations: MBL, mannose binding lectin; sp-A, sp-D, lung surfactant proteins A and D; CL-43, bovine serum collectin-43; CRD, carbohydrate recognition domain; MASP, mannose binding lectin-associated serine protease; PBS, phosphate buffered saline; BSA, bovine serum albumin; FSW, sterile filtered seawater; EDTA, ethylenediamine tetraacetic acid; galNAc, *N*-acetyl-D-galactosamine; gluNAc, *N*-acetyl-D-glucosamine; DTT, dithiothreitol; PVDF, polyvinyl difluoride; CAPS, cyclohexylaminopropane sulfonic acid; ODU, optical density units; SDS-PAGE, Sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

* Corresponding author. Tel.: +61-2-9850-8402; fax: +61-2-9850-8245.

E-mail address: draftos@rna.bio.mq.edu.au (D.A. Raftos)

0305-0491/00/$ - see front matter
PII: S0305-0491(99)00180-7

collagen-like domain that is typified by its repeating pattern of glycine-X-Y (Gly-X-Y) amino acid triplets (where X and Y can be any amino acid, frequently proline or hydroxyproline). A distinct neck region appears between the collagenous domain and the carboxy-terminal C-type carbohydrate recognition domain (CRD) (Holmskov et al., 1994; Epstein et al., 1996).

The distinctive domain structure of collectins is responsible for their characteristic patterns of oligomerization. Collectin polypeptides tend to associate into homotrimers, in which collagenous domains form collagen-like triple helices (Lu et al., 1993; Crouch et al., 1994; Holmskov et al., 1994). These helices are stabilized by hydrophobic interactions between leucines and valines in the neck region, as well as by interchain disulfide bonding between cysteine residues in the tail domain (Crouch et al., 1994; Hoppe and Reid, 1994). Most collectins also exhibit higher order structures (oligomers of collectin trimers) that are established by disulfide bonding between cysteines in the tail domain or by non-covalent interactions (Strang et al., 1986; Lu et al., 1993; Holmskov et al., 1994; Laursen et al., 1995).

Considerable interest has focused on collectins because of their ability to interact with components of the complement system (Lu, 1997). Complement is a major immunological effector mechanism in all classes of vertebrates (Dodds and Day, 1993). The initiation of proteolytic complement cascades has traditionally been associated with either antibody-based antigen recognition (the classical activation pathway) or spontaneous hydrolysis (the alternative activation pathway). However, it has recently been shown that complement can also be activated by the collectin, MBL, that is secreted during acute phase responses to infection (Holmskov et al., 1994; Epstein et al., 1996). MBL can recognize carbohydrates on the surface of pathogens and then activate the central complement component, C3, via a mannose binding lectin-associated serine protease (MASP) (Matsushita and Fujita, 1992; Tan et al., 1996; Thiel et al., 1997). MASP can either activate C3 directly or, more often, via the activation of the classical pathway C3 convertase.

Studies of a number of invertebrate species, which do not express immunoglobulin antibodies, indicate that the lectin-mediated complement pathway evolved before antibody-based complement activation mechanisms. Evolutionary homologs of C3 have been identified in both sea urchins (phylum *Echinodermata*) and tunicates (phylum *Urochordata*), which share common ancestors with the vertebrates (Smith et al., 1996; Al-Sharif et al., 1998; Nonaka et al., 1998). Sequence comparisons between invertebrate C3 homologs and their vertebrate counterparts suggest that the invertebrate C3 proteins have the capacity for proteolytic activation. It has also been shown that tunicates express serine proteases that are closely related to vertebrate MASPs (Ji et al., 1997), and complement Factor B homologs are evident in echinoderms (Smith et al., 1996). This confirms that critical components of lectin-mediated complement activation (C3 and MASP) are expressed in phylogenetically primitive deuterostomes. However, proteins analogous to MBL, which undertakes antigen recognition during lectin-mediated complement activation, have not yet been identified among invertebrates.

In this study, we characterize a calcium-dependent lectin from the tunicate, *Styela plicata*, which bears extensive physicochemical and amino acid sequence similarities with vertebrate collectins.

2. Experimental procedures

2.1. Tunicates

S. plicata were collected from boat moorings and sharkproof netting in Sydney Harbor with the permission of New South Wales Fisheries Service (Research permit # 95067). The tunicates were kept in 40 l aerated aquaria (15°C) until required for study. Hemolymph was harvested on ice from incisions in the buccal siphon. Exuding hemolymph was immediately centrifuged (2000 × *g*, 7 min, 4°C) and filtered (0.45 μm; Costar, Pleasanton, CA) before the addition of protease inhibitor (2 mM Pefabloc; Boehringer Mannheim, North Ryde, NSW).

2.2. Biotinylation of hemolymph

Hemolymph was concentrated to a final total protein content of 1 mg/ml with Centriprep 10 centrifugal concentrators (Amicon, Davers, MA). The concentrated hemolymph was then biotinylated with Immunopure NHS-LC-biotin (Pierce, Rockwood, IL) according to the manufacturer's instructions.

2.3. Carbohydrate binding assay

The ability of hemolymph components to bind defined carbohydrates was quantified using bovine serum albumin (BSA) neoglycoproteins (galactosyl, fucosyl, lactosyl, maltosyl, *N*-acetyl glucosamide, *N*-acetyl glucosaminide, *N*-acetyl galactosamide and *N*-acetyl galactosaminide; Sigma Chemicals, St Louis, MO) in a microplate spectrophotometric assay. Neoglycoproteins (100 μg/ml in 150mM phosphate buffered saline (PBS); pH 7.6) were bound to 96-well flat-bottomed RIA/EIA plates (Costar, Pleasanton, CA) by overnight incubation at room temperature. The plates were then washed extensively with PBS and blocked for 2 h with 5% w/v BSA (Sigma) in PBS. After further washing, wells were incubated with biotinylated hemolymph (20 μg/ml in sterile filtered seawater (FSW)) for 2 h (room temperature). Biotinylated hemolymph proteins that had bound to the plate were identified by incubation with streptavidin-alkaline phosphatase (1:5000 in PBS; Sigma) for 2 h followed by *p*-nitrophenyl phosphate (Sigma). Color development was quantified on a microplate reader at 405 nm. In some experiments, biotinylated hemolymph was incubated with potential inhibitors of binding (20 mM ethylenediamine tetraacetic acid (EDTA), 50 mM *N*-acetyl-D-galactosamine (galNAc), or 50 mM *N*-acetyl-D-glucosamine (gluNAc)) for 15 min before being added to plates. Fractions of hemolymph from galNAc affinity columns (see later) were tested for binding activity with or without the addition of 50 mM $CaCl_2$ or 50 mM $MgCl_2$. The statistical significance of differences between binding activities were determined by two-tailed Student's *t*-tests (Sokal and Rohlf, 1981).

2.4. Affinity purification of lectins from tunicate hemolymph

GalNac binding specificities were investigated further because previous studies have shown that opsonic activity in *S. plicata* is directed against galactose residues (unpublished data). Three milliliters of galNAc conjugated to 4% beaded agarose (Sigma Chemicals) were mixed with 50 ml non-biotinylated hemolymph or 1 ml biotinylated hemolymph overnight (4°C). The slurry was then decanted into a polypropylene chromatography column (BioRad, North Ryde, NSW) so that the beads could be washed with 50–100 ml FSW. Bound proteins were eluted with 3 ml EDTA (50 mM in 150 mM phosphate buffered saline; pH 7.2). Then, 200 μl fractions were collected.

2.5. Gel filtration

Fractions from the galNAc affinity column were subjected to gel filtration with a SMART high-performance liquid chromatography (HPLC) gel filtration system (Pharmacia, Uppsala, Sweden) using PBS containing 50 mM EDTA as the elution buffer. The column was calibrated with molecular mass standards in a 24–2000 kDa range (Sigma). Fifty microliter fractions were collected and subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis.

2.6. Electrophoresis

Fractions from the galNAc affinity and gel filtration columns were electrophoresed through 10% SDS-PAGE gels by the method of Laemmli (1970). When required, samples were reduced by the addition of 10 mg/ml w/v dithiothreitol (DTT) to sample buffers. Gels were silver stained using the method of Blum et al. (1987).

2.7. Two-dimensional electrophoresis

Affinity chromatography fractions of hemolymph were electrophoresed through precast 4–20% gradient gels (Ready Gels; BioRad) under non-reducing conditions. Lanes containing electrophoresed proteins were then excised and reduced by soaking for 10 min in 0.12 M Tris–HCl (pH 6.8) containing 10% v/v glycerol, 2% w/v SDS and 0.5% w/v DTT. The reduced gel slice was placed across the top of a 10% SDS-PAGE gel and electrophoresed before being silver stained (Blum et al., 1987).

2.8. Western blotting

Proteins were transferred from SDS-PAGE gels to polyvinyl difluoride (PVDF) membranes (Bio-Rad) using cyclohexylaminopropane sulfonic acid (CAPS) transfer buffer (10 mM CAPS (pH 11), 10% v/v methanol) according to the method of Aebersold et al. (1987). Blotted proteins were stained with Coomassie blue (0.5% w/v

Coomassie blue, 50% v/v methanol). Individual bands were excised, and destained in 50% v/v methanol before being subjected to either amino acid composition analysis, tryptic digestion or N-terminal amino acid sequencing.

2.9. Amino acid composition analysis

The amino acid compositions of bands excised from PVDF membranes were determined by Australian Proteome Analysis Facility (Macquarie University, NSW). Amino acid compositions were compared with data in the SwissProt data base using the AACOMPIDENT tool provided by the EXPASY Molecular Biology Server (http://expasy.hcuge.ch/) (Wilkins et al., 1996a,b). Searches were key-worded for the term 'lectin' and molecular mass tolerances were set at 43 ± 20 kDa. The statistical significance of differences in the amino acid compositions of different proteins were determined by χ^2 analysis (Moroney et al., 1951).

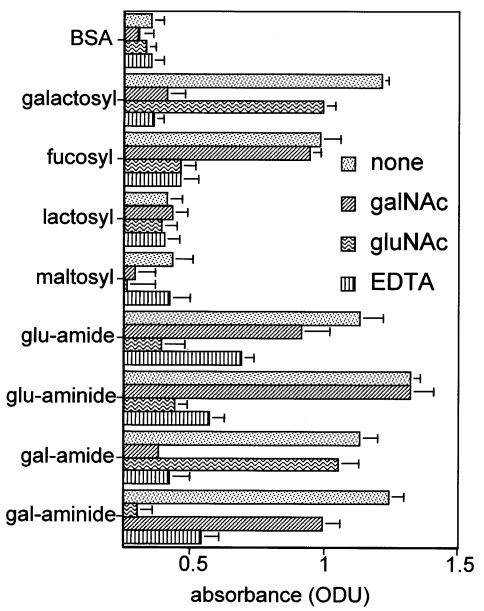

Fig. 1. Binding (absorbance as optical density units (ODU_{405})) of biotinylated hemolymph to BSA or a range of BSA neoglycoproteins that were immobilized on EIA/RIA plates. Hemolymph was incubated with BSA or neoglycoproteins, either without potential inhibitors of binding or in the presence of 20 mM EDTA, 50 mM gluNAc or 50 mM galNAc. Bars represent standard errors, $n = 6$. The neoglycoproteins used were galactosyl (BSA-galactose), fucosyl (BSA-fucosylamide), maltosyl (BSA-maltosyl), lactosyl (BSA-lactosyl), glu-amide(BSA-glucosamide), glu-aminide(BSA-*p*-aminophenyl-*N*-acetyl-β-D-glucosaminide), gal-amide (BSA-galactosamide), gal-aminide (BSA-*p*-aminophenyl-*N*-acetyl-β-D-galactosaminide).

2.10. Tryptic digestion

Proteins cut from PVDF membranes were reduced and alkylated before being digested overnight (37°C) with 1 μg trypsin (sequencing grade; Boehringer Mannheim, North Ryde, NSW) in 10 μl Tris–HCl (100 mM; pH 8.5) containing 1% acetyl glucoside. The digested PVDF membrane slice was placed in trifluoroacetic acid (TFA) (2% v/v in isopropanol) and sonicated for 10 min. The membrane was then further extracted with acetonitrile. The eluates from the PVDF membrane and the supernatant from the digest were pooled, dried in a centrifugal evaporator and resuspended in 0. 5%TFA. The pooled samples were then subjected to reverse-phase HPLC on a SMART HPLC system fitted with a Sephasil C8 column (Pharmacia) to isolate individual peptides for sequencing.

2.11. Collagenase digestion

Fifty microliter aliquots of galNAc affinity chromatography fractions were digested for various periods in 50 μl collagenase (Sigma Chemicals; 7.17 U in 10 mM KCl, 15 mM $MgCl_2$, 10 mM $CaCl_2$ and 10 mM Tris–HCl; pH 7.5). Collagenase-treated samples were resuspended in SDS-PAGE reducing sample buffer and electrophoresed through 12% SDS-PAGE gels.

2.12. Protein sequencing

Sequencing of PVDF immobilized proteins or HPLC fractions from tryptic digests was performed by the Macquarie University Center for Analytical Biochemistry using an Edman Sequenator (HP G1000 A Protein Sequencer; Hewlett Packard, Palo Alto, CA) with routine 3.1 and 3.1 PVDF chemistries.

3. Results

3.1. Carbohydrate binding activity of hemolymph

The ability of biotinylated hemolymph to bind various neoglycoproteins is shown in Fig. 1. Significantly enhanced binding ($P < 0.05$ relative to control wells containing immobilized BSA with no carbohydrate) was evident when hemolymph was incubated in wells containing galactosyl, fu-

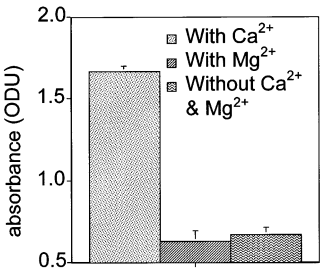

Fig. 2. Binding to immobilized BSA-*N*-acetyl galactosamide (absorbance as ODU_{405}) of a fraction of biotinylated hemolymph isolated by galNAc affinity chromatography. The fraction was incubated with BSA-*N*-acetyl galactosamide without added divalent cations (no Ca^{2+} or Mg^{2+}) or in the presence of 50 mM Ca^{2+} or 50 mM Mg^{2+}. Bars = standard errors, $n = 6$.

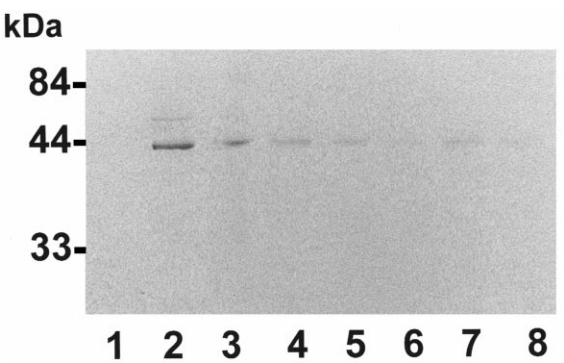

Fig. 3. Silver-stained reducing SDS-PAGE of the first eight fractions (lanes 1–8) eluted from a galNAc affinity chromatography column. The position of molecular mass markers (kDa) is shown on the left.

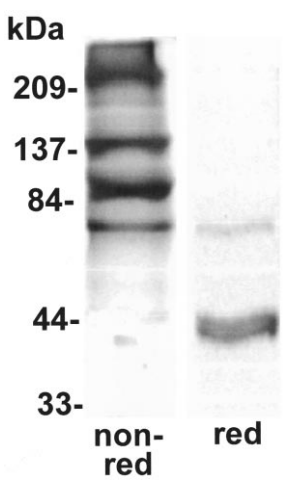

Fig. 4. galNAc affinity purified hemolymph proteins subjected to SDS-PAGE under either reducing (red) or non-reducing (non-red) conditions. The position of molecular mass markers (kDa) is shown on the left.

cosyl, *N*-acetyl glucosamide, *N*-acetyl glucosaminide, *N*-acetyl galactosamide and *N*-acetyl galactosaminide neoglycoproteins. No significant binding ($P > 0.05$ compared with BSA controls) was detected in wells containing lactosyl or maltosyl neoglycoproteins. In all cases, the binding of hemolymph to neoglycoproteins was significantly inhibited by pre-incubating hemolymph with 20 mM EDTA.

Pre-incubation of hemolymph with soluble *N*-acetyl-D-galactosamine inhibited the ability of hemolymph proteins to bind galactose, *N*-acetyl galactosamide and *N*-acetyl galactosaminide neoglycoproteins, but did not significantly affect binding to fucose, *N*-acetyl glucosamide, *N*-acetyl glucosaminide neoglycoproteins ($P > 0.05$). Conversely, soluble *N*-acetyl-D-glucosamine inhibited the binding of hemolymph proteins to fucose, *N*-acetyl glucosamide and *N*-acetyl glucosaminide, but not to galactose, *N*-acetyl galactosamide and *N*-acetyl galactosaminide.

Fig. 2 shows that a fraction of biotinylated hemolymph isolated by galNAc affinity chromatography could not bind wells coated with *N*-acetyl galactosamide-BSA without the addition of divalent cations to quench the effects of the EDTA that was used to elute protein from the affinity column. The *N*-acetyl galactosamide-BSA binding activity of the fraction was restored by co-incubation with 50 mM Ca^{2+}, but not 50 mM Mg^{2+}.

3.2. SDS-PAGE of galNAc affinity fractions

Under reducing conditions, SDS-PAGE showed that fractions 2–8 from a galNAc affinity column contained a 43 kDa polypeptide (Fig. 3). The highest titer of this polypeptide was evident in fraction 2. A less intense band at 65 kDa was evident in fractions 2 and 3, but not fractions 4–8 (Fig. 3).

Under non-reducing conditions, SDS-PAGE of galNAc fractions resolved proteins at 43, 65, 90, 140 and approx. 260 kDa (Fig. 4). Larger proteins (> 260 kDa) were also evident, but could not be efficiently resolved electrophoretically to determine their apparent molecular masses.

Two-dimensional electrophoresis indicated that the 90, 140 and 260 kDa non-reduced proteins (Fig. 5) could be reduced to 43 kDa. Reduction of the 65 kDa band did not alter its apparent molecular mass.

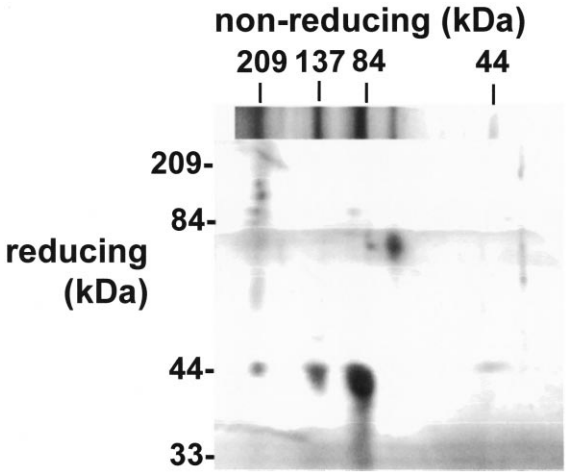

Fig. 5. Two-dimensional SDS-PAGE (non-reducing conditions, left to right; reducing conditions, top to bottom) of proteins isolated from hemolymph by galNAc affinity chromatography. A lane electrophoresed only in the first dimension (non-reducing) is also shown. The position of molecular mass markers (kDa) is shown at the top (non-reducing) and left (reducing).

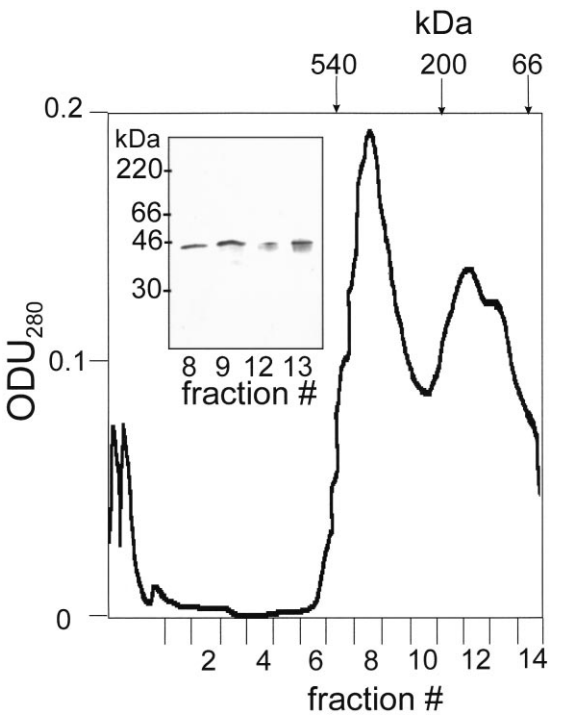

Fig. 6. Gel filtration of a fraction isolated from hemolymph by galNAc affinity chromatography. The elution times for molecular mass standards are shown by arrows. Inset: Reducing SDS-page of gel filtration fractions 8, 9, 12 and 13. The position of molecular mass markers (kDa) is shown on the left.

3.3. Gel filtration chromatography of galNAc affinity fractions

Fig. 6 shows that two protein peaks were evident when galNAc affinity fractions of hemolymph were subjected to gel filtration chromatography. The galNAc fractions used for gel filtration contained the 43 kDa reduced protein, but not the 65 kDa protein detected in some affinity fractions (see Fig. 3). The peaks detected in gel filtration corresponded to apparent molecular masses of approximately 400 and 130 kDa, with the 400 kDa peak being the most prominent. A shoulder also appears on the 130 kDa peak, suggesting the possibility of a third species. Reducing SDS-PAGE of gel filtration fractions containing the peaks and the shoulder resolved a single polypeptide at 43 kDa (Fig. 6).

3.4. Amino acid composition analysis

Comparison of the amino acid composition of the non-reduced 260 kDa tunicate protein from galNAc affinity chromatography with the SwissProt data base using the AACOMPIDENT algorithm returned 16 collectin polypeptides (conglutinin, MBL, sp-A, sp-D, CL-43) within the top 20 matches (Table 1). Bovine conglutinin had the highest similarity to the tunicate protein.

Table 2 shows the relative amino acid compositions of the reduced 43 kDa protein and the non-reduced 90 and 260 kDa proteins isolated by galNAc affinity chromatography. χ^2 analyses revealed no significant differences ($P > 0.05$) between the amino acid compositions of the three tunicate proteins. All three tunicate proteins exhibited high glycine contents (16–20%) that were similar to those of the vertebrate collectins (MBL-A, conglutinin and sp-D). This high glycine content differed substantially from that of a presumably unrelated protein (BSA) that was used to calibrate the AACOMPIDENT analysis. All three tunicate proteins also incorporated hydroxyproline, as did all of the collectins.

3.5. Collagenase digests

SDS-PAGE gels of a galNAc affinity fraction after collagenase digestion revealed a predominant fragment of approximately 26 kDa (Fig. 7). The band in lane 5 (1 h digestion) appears to be of slightly higher molecular mass than the bands

Table 1
The first 20 matches (in rank order) from an AACOMPIDENT search comparing the amino acid composition of the 260 kDa non-reduced form of the tunicate lectin with all lectins in the SwissProt database[a]

Rank	Score	MW	Description
1	107	35 749	Conglutinin*
2	113	33 398	Fucose-specific lectin
3	113	35 472	Sp-D, human*
4	122	24 473	Sp-A, pig*
5	124	35 651	Sp-D, mouse*
6	125	31 362	CL-43*
7	132	35 482	Sp-D, rat*
8	133	24 170	Sp-A, human*
9	141	35 115	Sp-D, bovine*
10	153	24 562	Sp-A, rabbit*
11	155	23 526	MBL-A, rat*
12	159	24 104	Sp-A, mouse*
13	163	24 396	Sp-A, dog*
14	170	64 004	E-Selectin, dog
15	172	23 605	MBL-A, mouse*
16	176	24 189	Sp-A, rat*
17	179	27 916	Myxobacterial hemagglutinin
18	180	50 243	E-Selectin, pig
19	181	24 040	MBL-C, rat*
20	187	23 967	MBL-C, mouse*

[a] The search was limited to lectins with MW 43 ± 20 kDa. Collectins are identified by *. The 'scores' indicated in the table represent comparisons with the tunicate collectin calculated as the sum of the square of differences in the molar percentage of amino acids.

in lanes 6 (6 h digestion) and 7 (overnight digestion), suggesting that the collagenous domain was incompletely digested in lane 5. The 26 kDa band could not be detected in non-digested samples or in samples containing collagenase alone. The 65 kDa protein detected in some affinity purified preparations (see Fig. 4) was not digested by collagenase (data not shown).

3.6. Amino acid sequencing

Sequences for the first 31 N-terminal amino acids of the reduced 43 kDa protein, and for an internal peptide generated by tryptic digestion, are shown in Fig. 8. N-terminal sequence tags (six amino acids) of the 90 and 260 kDa non-reduced tunicate proteins from galNAc affinity chromatography showed identical sequences to that of the 43kDa reduced protein (data not shown). The 43 kDa tunicate protein has a 19 amino acid N-terminal region (tail domain), including a single cysteine at residue 5. The tail domain abuts a series of Gly-X-Y triplet repeats, where X and Y are represented by a number of different amino acids, including prolines at residues 21 and 30. The internal fragment also shows a series of Gly-X-Y triplet repeats.

An alignment between the 43 kDa tunicate protein and a number of vertebrate collectins is also shown in Fig. 8. The alignment, which inserted a single amino acid gap in the sequences of the tunicate lectin and rat MBL-A, reveals 30% amino acid identity between the tunicate protein and rat MBL-A, and 27% identity between CL-43 and the tunicate protein. The cysteine at residue 5 of the tunicate lectin aligns precisely with a conserved cysteine in rat MBL-A.

4. Discussion

We have identified a calcium-dependent lectin from tunicates that bears numerous similarities to vertebrate collectins. The tunicate protein was purified from hemolymph by affinity chromatography using a galNAc affinity matrix. The choice of that carbohydrate matrix was based on data from carbohydrate binding assays. Biotinylated hemolymph was shown to bind glucose, galactose and fucose based neoglycoproteins. In all cases, this binding was abrogated by EDTA, suggesting that interactions were divalent cation dependent. It was subsequently demonstrated that proteins isolated by galNAc affinity chromatography were specifically calcium dependent.

The differential ability of soluble carbohydrates (galNAc or gluNAc) to inhibit binding indicated that at least two different carbohydrate binding (lectin) specificities were present in hemolymph. GluNAc inhibited binding to glucose- and fucose-based neoglycoproteins, whereas galNAc specifically inhibited binding to galactose-based moieties. In this study, we have further investigated the galactose binding specificity.

Under reducing conditions, the major polypeptide evident in galNAc affinity chromatography fractions of hemolymph had an apparent molecular mass of 43 kDa. This is similar to the apparent molecular masses of vertebrate collectin polypeptides (28–48 kDa), particularly conglutinin (48 kDa), sp-D (44 kDa) and CL-43 (43 kDa) (Lee et al., 1991; Lu et al., 1992). A minor band at 65 kDa was also evident in some galNAc fractions. We have not further characterized this protein.

Table 2
Amino acid compositions (mol%) of the reduced 43 kDa tunicate lectin, and its 90 and 260 kDa non-reduced forms[a]

Amino acid	Tunicate			BSA	Rat MBL-A (Oka et al., 1988)	Conglutinin (Davis and Lachmann, 1984)	Human sp-D (Lu et al., 1992)
	43 kDa	90 kDa	260 kDa				
Asx	9.5	7.4	4.3	9.5	9.5	8.8	6.9
Glx	8.9	9.7	13.6	12.9	12.7	11.8	13.3
Ser	12.1	11.5	11.3	5.7	7.6	6.2	5.3
His	2	0.7	0.6	1.2	2.3	1	6
Gly	16.6	20.5	20.1	4.3	13.2	19	23.4
Thr	6.1	5.8	6	6.8	6.8	4.4	2.5
Ala	8.7	8.8	7.7	9.6	7.3	8.3	9.6
Pro	5.4	4.3	6.1	6.4	5.4	9	6.3
Tyr	2.1	1.8	1.8	1.9	2.1	1.5	1.2
Arg	6.2	6.9	5.6	5.2	4.2	4.1	4.2
Val	5.6	5.7	5.8	7.4	5.6	4.3	4.8
Met	0.8	1.3	1.3	1	1.6	2	1.7
Ile	4.3	4	4.8	3.1	3.4	2.3	1.5
Leu	4.6	4.7	4.6	11	7.2	5	4.4
Phe	3.4	4.4	4.5	5.7	3.8	2.7	2
Lys	3.2	2.5	1.7	8.2	7.7	5.9	5.2
OH-Pro	2.4	0.5	0.5	0	1.1	3.7	3.2

[a] The amino acid compositions of BSA and a variety of mammalian collectins are also shown.

Non-reducing SDS-PAGE of the affinity purified tunicate lectin revealed numerous proteins with apparent molecular masses that were consistent with different multimerization states of the 43 kDa reduced polypeptide. The putative multimers had apparent molecular masses equivalent to the expected size of dimers (90 kDa), trimers (140 kDa), hexamers (260 kDa) and higher order oligomers (> 260 kDa). Two-dimensional SDS-PAGE confirmed that the 90, 140 and 260 kDa molecules were homo-oligomers of the reduced 43 kDa polypeptide. Such patterns of oligomerization are typical of vertebrate collectins (Laursen et al., 1995). Under non-reducing SDS-PAGE, sp-D, conglutinin and MBL appear as trimers, hexamers and higher order oligomers (Holmskov et al., 1995; Laursen et al., 1995). While none of the vertebrate collectins normally appear as dimers in non-reducing SDS-PAGE, both MBL and conglutinin occur as tetramers and pentamers, as well as trimers and hexamers (Crouch et al., 1994; Laursen et al., 1995). This suggests that the multimerization of collectins detected by SDS-PAGE need not be restricted to collagenous trimers and oligomers of those trimers.

Gel filtration of galNAc fractions indicated that the predominant native form of the tunicate lectin is a 400 kDa protein, which migrates at 43 kDa when examined by reducing SDS-PAGE. This suggests that most of the tunicate lectin normally appears as a nonamer. Again, this pattern is consistent with vertebrate collectins. Most native collectins appear as oligomers of collagenous trimers (Lu, 1997). For instance, human MBL normally occurs as an oligomer of six collagenous trimers (18 individual polypeptides) when isolated by gel filtration (Holmskov et al., 1994).

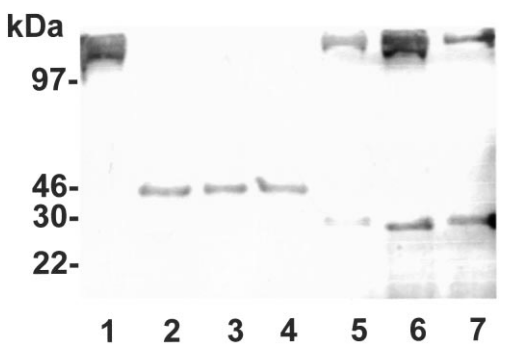

Fig. 7. Reducing SDS-PAGE of proteins isolated by galNAc affinity chromatography and then digested with collagenase. Lane 1, collagenase alone; lanes 2–4, galNAc fraction incubated for 1 h (lane 2), 6 h (lane 3) and overnight (lane 4) in buffer without collagenase; lanes 5–7, galNAc fractions incubated for 1 h (lane 5), 6 h (lane 6) and overnight (lane 7) in the presence of collagenase. The position of molecular mass markers (kDa) is shown on the left.

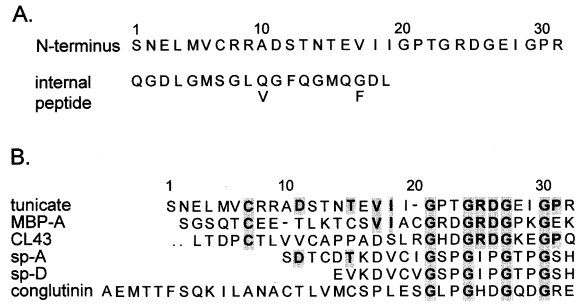

Fig. 8. (A) Amino acid sequences of the N-terminus and an internal tryptic peptide from the 43 kDa tunicate protein isolated from hemolymph by galNAc affinity chromatography. (B) Sequence alignments of the 43 kDa tunicate protein with a variety of collectins (rat MBL-A, human sp-A and sp-D, CL-43 and bovine conglutinin). Residues conserved between the tunicate protein and at least one of the collectins are highlighted.

Amino acid composition analysis of the tunicate lectin also revealed obvious similarities to collectins. The reduced 43 kDa tunicate polypeptide, and the 90 and 260 kDa non-reduced oligomers from SDS-PAGE, had glycine contents of approximately 20%. This high proportion of glycine, relative to presumably unrelated proteins such as BSA, is consistent with that found in vertebrate collectins (9–24%) and is indicative of collagen-like domains in which glycines appear at every third residue. The existence of collagenous regions in the tunicate lectin is further supported by the appearance of hydroxyproline, an amino acid that is characteristic of collagenous proteins including vertebrate collectins.

The inclusion of a collagenous domain in the tunicate lectin was confirmed by terminal and internal amino acid sequencing. N-Terminal sequence revealed a 19-amino-acid tail domain preceding a typical collagenous domain with Gly-X-Y repeats. Moreover, proline, which frequently occurs in the X or Y position of collagenous triplets in vertebrate collectins, appears in two of the first four Gly-X-Y repeats of the tunicate lectin. Gly-X-Y repeats were also evident in an internal peptide generated by tryptic digestion of the tunicate lectin.

Collagenase digestion of the tunicate lectin yielded a non-digested fragment of 26 kDa. This indicates that the digested collagenous domain is approximately 17 kDa. Given the calcium-dependent carbohydrate binding activity of the tunicate lectin, it is likely that the 26 kDa fragment remaining after collagenase digestion incorporates a carbohydrate binding domain (CRD) equivalent to that of vertebrate collectins. In bovine conglutinin, two fragments of 24 and 21 kDa, which incorporate the neck and CRD domains, remain after collagenase digestion (Davis and Lachmann, 1984).

Alignments of the tunicate lectin's N-terminal sequence with a number of collectins showed moderate levels of amino acid homology in the tail domain. Thirty percent amino acid identity was evident over a 30-amino-acid overlap between the tunicate lectin and rat MBL-A, and 27% identity was evident between CL-43 and the tunicate lectin. N-Terminal sequencing also revealed that the tail domain of the tunicate lectin (19 amino acids) is of very similar size to that of rat MBL-A (18 amino acid) and contains a cysteine residue (residue 5) in an identical position to one of the cysteines in the tail of rat MBL-A. Multiple cysteines in the tail domains of collectins are thought to be responsible for the oligomerization of trimers into higher order oligomers, and for stabilizing trimers via intra-chain disulfide bonds. The presence of a single cysteine in the tail of the tunicate lectin may explain why dimers, as well as timers and oligomers of trimers, appear when the tunicate lectin is analyzed by non-reducing SDS-PAGE. Crouch et al. (1994) suggested that the single cysteine found in the tail of rat sp-D predisposed that collectin to form collagenous trimers in which disulfide-bonded dimers interact non-covalently with an individual monomer.

Given the lack of more extensive sequence data, we cannot yet unequivocally conclude that the tunicate lectin is an evolutionary homolog of vertebrate collectins. However, the structural and amino-terminal sequence similarities between the tunicate lectin and vertebrate collectins implies that the tunicate protein may be functionally analogous to vertebrate collectins. The oligomeric nature of the tunicate lectin suggests that it may utilize mechanisms of antigen discrimination that are analogous to those of collectins (Epstein et al., 1996; Lu, 1997). The clustering of CRD derived from the oligomeric structure of defensive collectins such as MBL appears to confer their high affinity for 'non-self', as opposed to 'self', carbohydrate antigens (Epstein et al., 1996).

The appearance of a collagen domain in the tunicate lectin also suggests that it might have the capacity to initiate a lectin-mediated complement

pathway. Vertebrate MASPs, which proteolytically activate C3, are known to specifically associate with collagen domains of MBL (Matsushita et al., 1995). Moreover, MASP and C3 homologs have already been identified in tunicates (Ji et al., 1997; Nonaka et al., 1998). To confirm that a complement pathway which can be activated by collectin-like proteins exists in tunicates, we are currently testing the functional relationship between the collagenous lectin identified in this study and tunicate C3 homologs.

Acknowledgements

This study was supported by the Australian Research Council (grant # A97-942) and the Macquarie University Internal Research Grants Scheme. The research was facilitated by access to the Australian Proteome Analysis Facility established under the Australian Government's Major National Research Facilities Program.

References

Aebersold, R.H., Leavitt, J., Saavedra, R.A., Hood, L.E., Kent, S.B.H., 1987. Internal amino acid sequence analysis of proteins separated by one or two dimensional gel electrophoresis after in situ digestion on nitrocellulose. Proc. Natl. Acad. Sci. USA 84, 6970–6974.

Al-Sharif, W.Z., Sunyer, J.O., Lambris, J.D., Smith, L.C., 1998. Sea urchin coelomocytes specifically express a homologue of the complement component C3. J. Immunol. 160, 2983–2997.

Blum, H., Beier, H., Gross, H.J., 1987. Improved staining of plant proteins, RNA and DNA in polyacrylamide gels. Electrophoresis 8, 93–99.

Crouch, E., Persson, A., Chang, D., Heuser, J., 1994. Molecular structure of pulmonary surfactant protein D (SP-D). J. Biol. Chem. 269, 17311–17319.

Davis, A.E., Lachmann, P.J., 1984. Bovine conglutinin is a collagen-like protein. Biochemistry 23, 2139–2144.

Dodds, A.W., Day, A.J., 1993. In: Whaley, K., Loos, M., Weiler, J.M. (Eds.), Complement in Health and Disease. Kluwer, Boston, pp. 39–88.

Epstein, J., Eichbaum, Q., Sheriff, S., Ezekowitz, R.A.B., 1996. The collectins in innate immunity. Curr. Opin. Immunol. 8, 29–35.

Holmskov, U., Malhotra, R., Sim, R.B., Jensenius, J.C., 1994. Collectins: collagenous C-type lectins of the innate immune defense system. Immunol. Today 15, 67–74.

Holmskov, U., Laursen, S.B., Malhotra, R., Wiedemann, H., Timpl, R., Stuart, G.R., Tornoe, H., Madsen, P.S., Reid, K.B.M., Jensenius, J.C., 1995. Comparative study of the structural and functional properties of a bovine plasma C-type lectin, collectin-43, with other collectins. Biochem. J. 305, 889–896.

Hoppe, H.J., Reid, K.B.M., 1994. Collectins — soluble proteins containing collagenous regions and lectin domains — and their roles in innate immunity. Protein Sci. 3, 1143–1158.

Ji, X., Azumi, K., Sasaki, M., Nonaka, M., 1997. Ancient origin of the complement lectin pathway revealed by molecular cloning of mannan binding-protein-associated serine protease from a urochordate, the Japanese ascidian, *Halocynthia roretzi*. Proc. Natl. Acad. Sci. USA 94, 6340–6345.

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

Laursen, S.B., Hedemand, J.E., Thiel, S., Willis, A.C., Skriver, E., Madsen, P.S., Jensenius, J.C., 1995. Collectin in a non-mammalian species: isolation and characterisation of mannan binding protein (MBP) from chicken serum. Glycobiology 5, 553–561.

Lee, Y.-M., Leiby, K.R., Allar, J., Paris, K., Lerch, B., Okarma, T.B., 1991. Primary structure of bovine conglutinin, a member of the C-type animal lectin family. J. Biol. Chem. 266, 2715–2723.

Lu, J.H., 1997. Collectins — collectors of microorganisms for the innate immune system. Bioessays 19, 509–518.

Lu, J., Willis, A.C., Reid, K.B.M., 1992. Purification, characterisation and cDNA cloning of human lung surfactant protein D. Biochem. J. 284, 795–802.

Lu, J., Wiedeman, H., Holmskov, U., Thiel, S., Timpl, R., Reid, K.B.M., 1993. Structural similarity between lung surfactant protein D and conglutinin. Eur. J. Biochem. 215, 793–799.

Matsushita, M., Fujita, T., 1992. Activation of the classical complement pathway by mannose-binding protein in association with a novel C1s-like serine protease. J. Exp. Med. 176, 1497–1502.

Matsushita, M., Ezekowitz, R.A.B., Fujita, T., 1995. The gly-54 asp allelic form of human mannose-binding protein (BP) fails to bind MBL-associated serine protease. Biochem. J. 311, 1021–1023.

Moroney, M.J., 1951. Facts from Figures. Penguin Books, Reading.

Nonaka, M., Azumi, K., Ji, X., Namikawa-Yamada, C., Sasaki, M., Saiga, H., Dodds, A.W., Sekine, H., Homma, M.K., Matsushita, M., Endo, Y., Fujita, T., 1998. Opsonic complement component C3 in the solitary ascidian, *Halocynthia roretzi*. J. Immunol. 162, 387–391.

Oka, S., Ikeda, K., Kawasaka, T., Yamashina, I., 1988. Isolation and characterisation of two distinct mannan-binding proteins from rat serum. Arch. Biochem. Biophys. 260, 257–266.

Smith, L.C., Chang, L., Britten, R.J., Davidson, E.H., 1996. Sea urchin genes expressed in activated coelomocytes are identified by expressed sequence tags. J. Immunol. 156, 593–602.

Sokal, R.R., Rohlf, F.J., 1981. Biometry. W.H. Freeman, London.

Strang, C.J., Slayter, H.S., Lachmann, P.J., David, A.E., 1986. Ultrastructure and composition of bovine conglutinin. Biochem. J. 234, 381–389.

Tan, S.M., Chung, M.C.M., Kon, O.L., Thiel, S., Lee, S.H., Lu, J., 1996. Improvement on the purification of mannan-binding lectin and demonstration of its Ca^{2+}-independent association with a C1s-like serine protease. Biochem. J. 319, 329–332.

Thiel, S., Vorup-Jensen, T., Stover, C.M., Schwaeble, W., Laursen, S.B., Poulsen, K., Willis, A.C., Eggleton, P., Hansen, S., Holmskov, U., Reid, K.B.M., Jensenius, J.C., 1997. A second serine protease associated with mannan-binding lectin that activates complement. Nature 386, 506–510.

Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J.C., Yan, J.X., Gooley, A.A., Hughes, G., Humphery-Smith, I., Williams, K.L., Hochstrasser, D.F., 1996a. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Bio/Technology 14, 61–65.

Wilkins, M.R., Ou, K., Appel, R.D., Sanchez, J.C., Yan, J.X., Golaz, O., Farnsworth, V., Cartier, P., Hochstrasser, D.F., Williams, K.L., Gooley, A.A., 1996b. Rapid protein identification using N-terminal 'sequence tag' and amino acid analysis. Biochem. Biophys. Res. Commun. 221, 609–613.