HCA and HML isolated from the red marine algae *Hypnea cervicornis* and *Hypnea musciformis* define a novel lectin family

CELSO S. NAGANO,[1] HENRI DEBRAY,[2] KYRIA S. NASCIMENTO,[3] VICENTE P.T. PINTO,[3] BENILDO S. CAVADA,[3] SILVANA SAKER-SAMPAIO,[4] WLADIMIR R.L. FARIAS,[4] ALEXANDRE H. SAMPAIO,[4] AND JUAN J. CALVETE[1]

[1]Instituto de Biomedicina de Valencia, Consejo Superior de Investigaciones Científicas (CSIC), E-46010 Valencia, Spain
[2]Laboratoire de Chimie Biologique et Unité Mixte de Recherche, Centre National de la Recherche Scientifique (UMR CNRS) 8576, Université des Sciences et Technologies de Lille (USTL), Villeneuve d'Ascq, France
[3]Departamento de Bioquímica e Biologia Molecular and [4]Laboratorio de Bioquímica Marinha, Departamento de Engenharia de Pesca, Universidade Federal do Ceará, Fortaleza, CE 60451-970, Brazil

(RECEIVED April 5, 2005; FINAL REVISION May 15, 2005; ACCEPTED May 22, 2005)

Abstract

HCA and HML represent lectins isolated from the red marine algae *Hypnea cervicornis* and *Hypnea musciformis*, respectively. Hemagglutination inhibition assays suggest that HML binds GalNAc/Gal substituted with a neutral sugar through 1–3, 1–4, or 1–2 linkages in O-linked mucin-type glycans, and Fuc(α1–6)GlcNAc of N-linked glycoproteins. The specificity of HCA includes the epitopes recognized by HML, although the glycoproteins inhibited distinctly HML and HCA. The agglutinating activity of HCA was inhibited by GalNAc, highlighting the different fine sugar epitope-recognizing specificity of each algal lectin. The primary structures of HCA (9193 ± 3 Da) and HML (9357 ± 1 Da) were determined by Edman degradation and tandem mass spectrometry of the N-terminally blocked fragments. Both lectins consist of a mixture of a 90-residue polypeptide containing seven intrachain disulfide bonds and two disulfide-bonded subunits generated by cleavage at the bond T^{50}–E^{51} (HCA) and R^{50}–E^{51} (HML). The amino acid sequences of HCA and HML display 55% sequence identity (80% similarity) between themselves, but do not show discernible sequence and cysteine spacing pattern similarities with any other known protein structure, indicating that HCA and HML belong to a novel lectin family. Alignment of the amino acid sequence of the two lectins revealed the existence of internal domain duplication, with residues 1–47 and 48–90 corresponding to the N- and C-terminal domains, respectively. The six conserved cysteines in each domain may form three intrachain cysteine linkages, and the unique cysteine residues of the N-terminal (Cys46) and the C-terminal (Cys71) domains may form an intersubunit disulfide bond.

Keywords: red marine algal lectins; novel protein family; *Hypnea cervicornis* lectin; *Hypnea musciformis* lectin; cysteine-rich proteins

The recognition of carbohydrates by proteins underlies key cellular processes, such as cell communication, host defense, fertilization, development, parasitic infection, and tumor metastasis. Lectins are the carbohydrate-binding proteins of non-immune origin found in all types of living organisms that decipher the glycocodes encoded in the structure of glycans attached to soluble and integral cell membrane glycoconjugates (Gabius and Gabius 1997). Mechanisms for sugar recognition in microorganisms, plants, and animals have evolved independently in diverse protein frameworks (Elgavish

Reprint requests to: Juan J. Calvete, Instituto de Biomedicina de Valencia, CSIC, Jaime Roig, 11, E-46010 Valencia, Spain; e-mail: jcalvete@ibv.csic.es; fax: +34-96-3690800.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.051498505.

and Shaanan 1997). The largest and best characterized lectin family is that from terrestrial plants, and accumulating evidence indicates that the vast majority of these lectins can be classified into four large and three small families of structurally and evolutionarily related proteins (Van Damme et al. 1998; consult also the 3D Lectin Database at http://webenligne.cermav.cnrs.fr/lectines/). The molecular structure and functional features of an increasing number of terrestrial plant lectins have been reported (Loris 2002). However, and in marked contrast to higher land plant lectins, marine algal lectins have been isolated and characterized at a much lower pace since the first report of hemagglutinating activity in these organisms almost 40 years ago (Boyd et al. 1966). Moreover, to date, biochemical and structural information on algal lectins is scarce and from only a few species, and hence the functional and phylogenetic classification of these lectins remains obscure. The available structural information indicates the existence of different carbohydrate-binding proteins in the marine algae investigated, including the green algae *Enteromorpha prolifera* (Ambrosio et al. 2003) and *Ulva pertusa* (Wang et al. 2004), and the red marine algae *Bryothamnion triquetrum* (Calvete et al. 2000), *Hypnea japonica* (Hori et al. 2000), *Hypnea musciformis* (Nagano et al. 2002), and species of the *Eucheuma* (Kawakubo et al. 1999) and *Ptilota* (Sampaio et al. 1998) genera. Moreover, the complete amino acid sequences of only three algal lectins have been determined (Calvete et al. 2000; Hori et al. 2000; Wang et al. 2004). These lectins do not display sequence similarity to any known plant lectin. Here we report the biochemical characterization and primary structures of two lectins (HCA and HML) isolated from the Brazilian red marine algae *Hypnea cervicornis* and *Hypnea musciformis*. HCA and HML are homologous proteins that belong to a novel protein family, showing the existence of structural diversity among the lectins of closely related *Hypnea* species living in distant ecosystems, namely, the Pacific coast of Japan and the Atlantic coast of Brazil.

Results and Discussion

Purification and hemagglutination activity of the H. cervicornis *and* H. musciformis *lectins*

Agglutinating activity against native and trypsin-treated rabbit erythrocytes has been reported in the aqueous extracts of the red marine algae *H. cervicornis* and *H. musciformis* (Ainouz and Sampaio 1991). The agglutinins, termed HCA (*Hypnea cervicornis* agglutinin) and HML (*Hypnea musciformis* lectin), were isolated by ammonium sulfate precipitation, ion-exchange chromatography, and reverse-phase HPLC.

The agglutination of native and trypsin-treated rabbit red blood cells by HML was not inhibited by any of the monosaccharides or disaccharides tested (even at 75 mM concentration) or by the polysaccharides carrageenan and fucoidan at a concentration of 2.5 mg/mL (Table 1). This result, which is in line with previous reports indicating that in general algal lectins do not show affinity for simple sugars, but exhibit binding activity for complex oligosaccharides and glycoproteins (Rogers and Hori 1993; Calvete et al. 2000; Nagano et al. 2002), hampered a precise assignment of the saccharide specificity of HML. On the other hand, this was not the case with HCA, as *N*-acetyl-D-galactosamine at 9.3 mM inhibited the agglutination of native rabbit erythrocytes induced by this lectin (Table 1).

The glycan-recognizing specificity of HCA and HML was investigated through hemagglutination inhibition

Table 1. *Inhibition by monosaccharides, polysaccharides, and glycoproteins of the hemagglutinating activity of the lectins from* Hypnea musciformis *(HML) and* Hypnea cervicornis *(HCA)*

	HML	HCA
Mono- and disaccharides	(mM)	(mM)
D-Glucose	>75	>75
D-Mannose	>75	>75
D-Galactose	>75	>75
Methyl-α-D-galactopyranoside	>75	>75
L-Fucose	>75	>75
N-Acetyl D-galactosamine	>75	9.3
N-Acetyl D-glucosamine	>75	>75
Lactulose	>75	>75
Polysaccharides	(μg/mL)	(μg/mL)
Carrageenan	>2500	>2500
Fucoidan	>2500	>2500
N-Glycoproteins	(μg/mL)	(μg/mL)
Human serotransferrin	>2500	156
Desialylated human serotransferrin	>2500	78
α1 acid glycoprotein	>2500	>2500
Desialylated α1 acid glycoprotein	312	2500
Human lactotransferrin	78.1	9.7
Desialylated human lactotransferrin	39	4.8
Hen ovomucoid	>2500	>2500
Hen ovalbumin	>2500	>2500
Porcine thyroglobulin	2.4	9.7
Desialylated porcine thyroglobulin	2.4	1.2
Bovine lactotransferrin	1250	9.7
Desialylated bovine lactotransferrin	78.1	4.8
Yeast mannan	>2500	>2500
O-Glycoproteins	(μg/mL)	(μg/mL)
Bovine fetuine	1250	156
Bovine asialofetuin	78	9.7
Bovine submaxillary mucin	0.6	0.6
Porcine stomach mucin	0.3	9.7
Ovine submaxillary mucin	39	4.8
Desialylated ovine submaxillary mucin	9.7	2.4

The titer and dilution (in between brackets) used were 64 (1/16) and 32 (1/8) for HML and HCA, respectively.

assays using an array of glycoproteins containing definite oligosaccharides. As shown in Table 1, the agglutinating activity of HML and HCA was efficiently inhibited by some glycoproteins bearing either complex type N-glycans (human and bovine lactotransferrin; porcine thyroglobulin) or O-glycans (porcine stomach mucin; ovine and bovine submaxillary mucins and their desialylated forms). However, the activity of HML and HCA could not be impaired by glycoproteins bearing high-mannose type N-glycans (yeast mannan), glycoproteins carrying both N-linked *N*-acetyllactosamine- and hybrid-type glycans (hen ovomucoid) or high-mannose type- and hybrid-type N-glycans (hen ovalbumin) (Table 1).

The best inhibitors of the *H. musciformis* lectin were the mucins from porcine stomach (0.3 μg/mL) and bovine submaxillar gland (0.6 μg/mL), and to a lesser extent (9.7 μg/mL) the desialylated ovine submaxillary mucin (Table 1). The porcine stomach mucin contains O-linked carbohydrate structures sharing the core 1 Galβ1–3GalNAc disaccharide, which can be substituted by *N*-acetyllactosamine branches terminated with fucose α1–2-Gal (human blood group H), GalNAcα1–3[Fuc α1–2]Gal (human blood group A), or GlcNAcα1–4-Gal at their nonreducing ends (Table 2). Tn (GalNAcα1-Ser/Thr) and T (Galβ(1–3)GalNAcα1-Ser/Thr) antigens are also present in the porcine stomach mucin (Van Halbeek et al. 1982; Zenteno et al. 1995; Karlsson et al. 1997). The bovine submaxillary mucin is a glycoprotein bearing at least 16 different structures (Savage et al. 1990, 1991; Chai et al. 1992). Of its oligosaccharides, 85% are acidic O-linked oligosaccharide chains, including a high density of sialyl Tn antigens and sialyl core 3 saccharide sequences (Table 2). The neutral O-linked glycans of bovine submaxillary mucin include the human blood groups A and H, and the core 3 determinants (Savage et al. 1990, 1991; Chai et al. 1992) (Table 2). On the other hand, the Tn antigen accounts for > 75% of the carbohydrate chains of the desialylated ovine submaxillary mucin (Table 2). As a whole, these data indicate that a preferred carbohydrate ligand of HML may be GalNAc/Gal substituted with a neutral sugar through 1–3, 1–4, or 1–2 linkages. Comparison of the blocking activities of the sialylated versus the desialylated ovine submaxillary mucin, and between the

Table 2. *Major saccharide determinants of* N- *and* O-*glycoproteins that inhibited the agglutinating activity of the lectins from* Hypnea musciformis *(HML) and* Hypnea cervicornis *(HCA)*

	Saccharide determinant	
N-Glycoproteins		
Porcine thyroglobulin; human and bovine lactoferrin		
	NeuAcα 2–6 Gal β 1–4 GlcNAc β 1–2 Man | α1–6 **Fuc α 1-6** | Man β 1–4 GlcNAc β 1–4 **GlcNAc β 1-ASN** / α 1-3 NeuAcα 2–6 Gal β 1–4 GlcNAc β 1–2 Man	
O-Glycoproteins		
Ovine submaxillary mucin	NeuAc α 2–6 GalNAc α 1Ser/Thr	(Sialyl Tn)
Desialylated ovine submaxillary mucin	GalNAc α 1Ser/Thr	(Tn antigen)
Porcine stomach mucin		
	GalNAc α 1Ser/Thr	
	Galβ 1–3 GalNAc α 1Ser/Thr	(T antigen)
	GlcNAc α 1–4-Gal	
	Fuc α 1–2 Gal-	(Blood group H)
	GalNAc α 1–3[Fucα 1–2] Gal-	(Blood group A)
Bovine submaxillary mucin		
	NeuAc α 2–6 GalNAc α 1Ser/Thr	
	NeuAc α 2–6[GlcNAc β 1–3] GalNAc α 1Ser/Thr	(sialyl core 3)
	GlcNAc β 1–3 GalNAC α 1Ser/Thr	(core 3)
	Fuc α 1–2 Gal-	
	GalNAc α 1–3 [Fuc α 1-2] Gal-	
Asialofetuin		
	Gal β 1–3 GalNAc α 1Ser/Thr	(core 1)

The putative epitope recognized by HML and HCA within *N*-glycoproteins is highlighted in boldface.

bovine submaxillary and the porcine stomach mucins, clearly indicated that the presence of α2–6-linked sialic acid impaired the blocking activity of the carbohydrates toward HML.

Among the N-glycoproteins tested, porcine thyroglobulin and its asialo form proved also to be good inhibitors of HML (Table 1). This glycoprotein exhibits a complex pattern of glycosylation. It bears two types of chains: oligomannose type (unit A) and *N*-acetyllactosamine type (unit B). Among the latter, the major N-glycans are mono- and disialylated α1–6-fucosylated biantennary-structures terminated with α2–6-linked sialic acid (Neu5Ac or Neu5Gc) on the Man α1–3 antennae. The Man α1–6 antennae show large heterogeneity. They can be terminated with Man, GlcNAc, or Gal, and the terminal Gal residue can be extended with Gal α1–3-, (Neu5Ac or Neu5Gc) α2–6-, or Neu5Ac α2–3-linked residues. Moreover, 3–0-sulfated Gal and 6–0-sulfated GlcNAc residues have also been reported (De Waard et al. 1991). Though this situation greatly complicates the assignment of possible saccharide determinants recognized by the lectin, the fact that human serotransferrin, a glycoprotein with diantennary and some minor triantennary *N*-acetyllactosamine-type glycans without any α1–6-linked core fucose residues (Spik et al. 1975), does not inhibit the agglutinating activity of HML (Table 1) suggests that the Fuc α1–6 GlcNAc core sequence of the di- and triantennary glycans of porcine thyroglobulin could represent an epitope recognized by HML (Table 2). Indeed, this sequence is more accessible to the lectin in 12% of the monosialylated diantennary glycans with the Man α1–6 branch ending with nonreducing Man or GlcNAc residues, than when the Man α1–6 branch is extended with *N*-acetyllactosamine or sialyl α2–6 *N*-acetyllactosamine sequences. In the latter case, the extended antenna can fold over the Man- or GlcNAc-terminated branch, thereby masking the Fucα1–6-GlcNAc determinant (Rademacher et al. 1986). Moreover, native or desialylated human lactotransferrin, which possess two α1–6-fucosylated diantennary *N*-acetyllactosamine-type glycans per molecule (Spik et al. 1982), were relatively good inhibitors of HML, although they were 32- and 16-fold weaker, respectively, than porcine thyroglobulin (Table 1), and desialylated bovine lactotransferrin, containing both oligomannose-type and heterogeneous *N*-acetyllactosamine-type N-linked glycans, 4% of which are α1–6-fucosylated diantennary chains (Codeville et al. 1992), was 32-fold less inhibitory than porcine thyroglobulin (Table 1).

The best inhibitor of the agglutinating activity of the *H. cervicornis* agglutinin (HCA) was the bovine submaxillary mucin, followed by the desialylated ovine submaxillary mucin, porcine stomach mucin, and asialofetuin (Table 1). These results suggested that, like HML, the HCA lectin may preferably bind to nonsialylated GalNAc/Gal substituted with a neutral sugar through 1–3, 1–4, or 1–2 linkages, though both lectins appear to exhibit distinct specificities. In line with this conclusion, HCA but not HML was inhibited by the monosaccharide GalNAc (Table 1). On the other hand, as was the case with HML, N-glycoproteins bearing α1–6-fucosylated *N*-acetyllactosamine-type glycans (bovine and human lactotransferrins, porcine thyroglobulin) were good antagonists of HCA (Table 1). However, these glycoproteins inhibited distinctly HML and HCA, further highlighting the different fine sugar epitope-recognizing specificity of each algal lectin.

Galactoside-specific lectins that are not related to HCA and HML were previously isolated from the red marine algae *Ptilota filicina* (PFL, 19.3 kDa) (Sampaio et al. 1998), *Ptilota serrata* (PSL, 18.4 kDa) (Sampaio et al. 1999), and *Ptilota plumosa* (PPL, 17.4 kDa) (Sampaio et al. 2002). PFL and PSL were inhibited by *N*-acetylgalactosamine, D-galactose, and their C1-nitrophenyl-(α or β) derivatives, and the presence of an acetamido group at C2 enhanced the sugar binding to the lectins. As described above for HCA and HML, porcine stomach mucin, and to a lesser extent bovine submaxillary mucin, were also potent inhibitors of PFL (4.8 and 310 μg/mL, respectively) and PSL (< 4.8 and 1250 μg/mL, respectively). Similar to the case with HCA and HML, elimination of sialic acid rendered the bovine mucin 4 and 65 times more inhibitory of the hemagglutinating activity of PFL and PSL, respectively. Additionally, neither sialoglycoproteins (fetuin, α1-acid glycoprotein, transferrin, lactotransferrin) nor high-mannose-type glycoproteins (ovomucoid, thyroglobulin, ovalbumin) inhibited the activity of PFL and PSL. On the other hand, the *P. plumosa* lectin exhibited human blood group B (Galα1–3[Fucα1–2]Gal)-agglutinating specificity, and this activity was inhibited by galactose, glucose, and their derivatives. However, all the glycoproteins tested failed to block the hemagglutinating activity of the lectin (Sampaio et al. 2002).

As a whole, the results indicate that related, though distinct, galactoside-binding activities have emerged in structurally unrelated lectins from different marine red alga species.

Biochemical characterization of the H. cervicornis *and* H. musciformis *lectins*

Reversed-phase HPLC analysis of the purified native lectins yielded homogeneous chromatographic peaks of apparent molecular masses by SDS-PAGE of 16.5 kDa (Fig. 1, lanes a,b). However, the MALDI-TOF masses of HCA and HML were, respectively, 9193 ± 3 Da and 9357 ± 1 Da (Fig. 2), indicating that these lectins exhibit anomalous electrophoretic mobility. On the other hand,

Figure 1. SDS-PAGE of the purified agglutinins. SDS-PAGE analysis of the lectins from the Brazilian red marine algae *H. cervicornis* (HCA) (lane *a*) native, (lane *c*) reduced and carbamidomethylated; and *H. musciformis* (HML) (lane *b*) native, (lane *d*) reduced and carbamidomethylated. (Lane *s*) Molecular mass standards.

SDS-PAGE under reducing conditions showed that both HCA and HML contained two disulfide-bonded polypeptide chains of apparent molecular masses of 8 and 4.5 kDa (Fig. 1, lanes c,d). The MALDI-TOF masses of native HCA and HML did not change upon incubation of the lectins with iodoacetamide under denaturing but non-reducing conditions, ruling out the presence of free sulfhydryl groups in their structures. Mass spectrometric analysis of the reversed-phase-separated fragments of reduced and carbamidomethylated (CM-) HCA showed ions at $m/z = 9990$ and 5643 (peak 2) and 4380 (peak 1) (Fig. 3), while CM-HML displayed ions at $m/z = 10{,}161$ and 5734 (peak 2) and 4460 (peak 1) (Fig. 3). As a whole, these data clearly indicated that HCA might consist of a mixture of a single-chain polypeptide containing 14 cysteine residues engaged in the formation of seven intrachain disulfide bonds [(9990 − 9193)/58 = 13.74 Cys] and two disulfide-bonded subunits with oxidized methionine residues after reduction and carbamidomethylation [(5643 + 4380) − 32 (2 Met-ox) = 9991 Da] containing seven (intra- + inter-) cysteine linkages. Similarly, MALDI-TOF mass spectra of CM-HML exhibited ions at $m/z = 10{,}161$, 5734, and 4460, which were interpreted as a mixture of a full-length polypeptide with 14 cysteine residues [(10,161 − 9357)/58 = 13.86 Cys] and the same molecule built by two subunits and containing two methionine sulfoxides [(5734 + 4460) − 32 (2 Met-ox) = 10,162 Da]. As judged by the SDS-PAGE analysis (Fig. 1), the major molecular species of both HCA and HML may consist of cleaved, disulfide-bonded subunits.

The primary structures of HCA and HML

Edman degradation of native HCA and HML showed the N-terminal sequences EAQFPVCCTTQWGLVSATGN CACGCMGG and EAQIPSCCSSRWGLVSVTGKCAC DCLGG, respectively. The molecular species detected by mass spectrometry upon reduction and carbamidomethylation of HCA and HML were isolated by reversed-phase HPLC and submitted to N-terminal sequencing. The single-chain species at $m/z = 9990$ (HCA) and 10,161 (HML), as well as the fragments of 5643 Da (HCA) and 5734 Da (HML), had blocked N termini, whereas the N-terminal sequences of the 4380-Da (HCA) and the 4460-Da (HML) subunits were identical to those of their parent native molecules. These results indicated that the 5.6/5.7 kDa (HCA/HML) and the 4.3/4.4 kDa (HCA/HML) fragments corresponded, respectively, to the N-terminal and C-terminal halves of the full-length, single-chain lectins. Determination of the complete amino acid sequences of HCA and HML was done by combination of Edman degradation of overlapping peptides obtained by proteolysis of the reduced and carbamidomethylated full-length lectins and their subunits and collision-induced fragmentation by MS/MS analysis of the N-terminal blocked chymotryptic peptides Q7 (HCA) and Q6 (HML) (Figs. 4,5). The product ion spectra of the doubly charged ion at $m/z = 1269.5$ (Q7 of HCA, $M + H^+$ 2537.06) and of the singly charged ion at $m/z = 779.3$ (Q6 of HML) produced sequence-specific b- and y-ion series from b2 to y1, from which their primary structures were interpreted as (240.1)ACTMDYFPVCCQ IVLGGGAY and (240.1)FCTL (Fig. 6A), respectively. The b2 ions at $m/z = 240.1$ were subfragmented using the $(MS)^3$ option of the QTrap linear ion trap spectrometer, yielding in each case the same sequence: ZK, where Z corresponded to a pyroglutamic acid residue (Fig. 6B).

Figure 2. MALDI-TOF mass spectrometry. Determination of the molecular mass of the native lectins isolated from the Brazilian red algae *H. cervicornis* (HCA) (*A*) and *H. musciformis* (HML) (*B*).

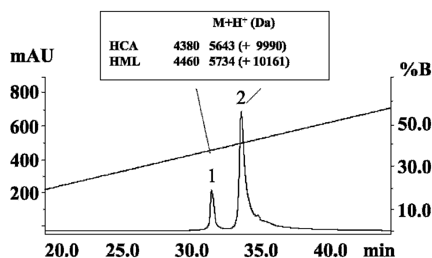

Figure 3. Reversed-phase HPLC. Separation of polypeptides after reduction and carbamidomethylation of HCA from *H. cervicornis*. A similar result was obtained for CM-HML from *H. musciformis*. The MALDI-TOF masses of the HCA and HML fragments recovered in each chromatographic peak are displayed in the box.

The amino acid sequences of HCA and HML each contains 90 residues and displays 55% sequence identity (80% similarity), including 14 conserved cysteine residues engaged in the formation of seven disulfide bonds. However, they exhibit neither discernible amino acid sequence similarity with, nor a cysteine spacing pattern found in any other known protein structure (see below), strongly indicating that HCA and HML belong to a novel protein (lectin) family. It is worth noting that the primary structures of HCA and HML are clearly different from those of the taxonomically related *Hypnea japonica* isolectins A1 and A2 (Hori et al. 2000). The latter resemble the evolutionarily more distant *Bryothamnion triquetrum* agglutinin (Calvete et al. 2000). Hence, our results highlight the occurrence of lectins of different algal species and the occurrence of agglutinins from different protein families isolated from species belonging to the same genus.

Structural features of HCA and HML

The high cysteine (disulfide bond) content of HCA and HML is an unusual feature of lectin structures. Among plant lectins, the only other known example of cysteine-rich proteins are the chitin-binding lectins, which are made up of hevein domains comprising ~40 residues, including eight conserved cysteine residues that are all involved in intrachain disulfide bonds (Cys3–Cys18, Cys12–Cys24, Cys17–Cys31, and Cys37–Cys41 in hevein, the rubber tree latex lectin), and a carbohydrate-binding site (Van Damme et al. 1998). Alignment of the amino acid sequences of HCA and HML against themselves revealed the existence of internal domain duplication (Fig. 7). Residues 1–47 and 48–90 of each lectin correspond to the N- and the C-terminal domains, respectively. Hence, generation of the two-subunit HML and HCA lectin species is accomplished by proteolytic cleavage at the peptide bond between residues 50 and 51 within the short polypeptide segment connecting the N- and the C-terminal domains. It is worth noting that each of these modules contains seven cysteine residues, six of which are conserved in position (Fig. 7). The cysteine spacing patterns of the N- and C-terminal domains of HCA and HML are $C_1(7)C_2C_3(14)C_4(1)C_5(9)C_6(5)\underline{C}$ and $C_{1'}(8)C_{2'}C_{3'}(12)$ $\underline{C}(1)C_{4'}\,(1)C_{5'}(10)C_{6'}$. Homologous cysteines have the same numbering, and the unique cysteine residues within each domain are underlined. Using the CysView program (Lenffer et al. 2004; available at http://research.i2r.a-star.edu.sg/CysView/) and the Disulphide Database (DSDBASE) (Vinayagam et al. 2004; at http://www.ncbs.res.in/~faculty/mini/dsdbase/dsdbase.html), no protein in the public available databanks showed a similar cysteine-pairing pattern or a similar disulfide bond connectivity. This result further strengthened our conclusion that HCA and HML truly belong to a novel protein family.

Though the pattern of disulfide bonding remains to be determined, we hypothesize that the six conserved cysteines in each domain may form three intrachain cysteine linkages and that the unique cysteine residues of the N-terminal (Cys46) and the C-terminal (Cys71) (Fig. 7) domains may form an intersubunit disulfide bridge.

Domain duplication is a general mechanism for enhancement/diversification of protein structure and function during evolution. However, whether the tandemly arranged domains of HML and HCA harbor independent carbohydrate-binding pockets or both contribute to the

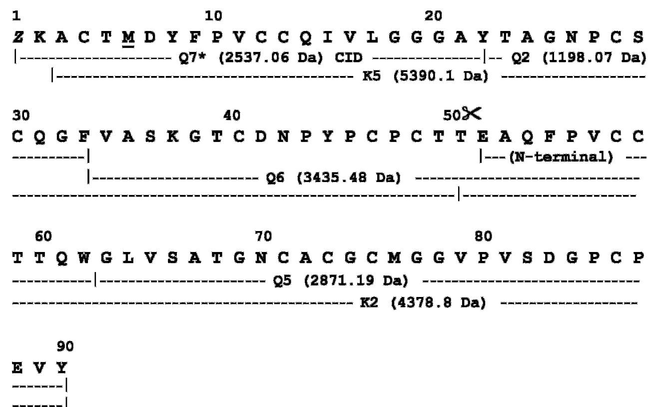

Figure 4. Amino acid sequence of HCA. The primary structures of HCA were determined by combination of Edman degradation of sets of overlapping peptides obtained by proteolysis of the reduced and carbamidomethylated full-length lectin and its reversed-phase HPLC fragments (isolated as in Fig. 3) with chymotrypsin (Q-) and endoproteinase Lys-C (K-), and by CID MS/MS analysis of the N-terminal blocked peptide Q7. Methionine residue at position 6 was oxidized in Q7. Proteolysis at the Thr50–Glu51 peptide bond, which generates the N-terminal sequence determined in the native two-chain HCA lectin and in its 4390-Da C-terminal fragment (identical to peptide K2), is indicated by scissors. (Z) Pyroglutamic acid.

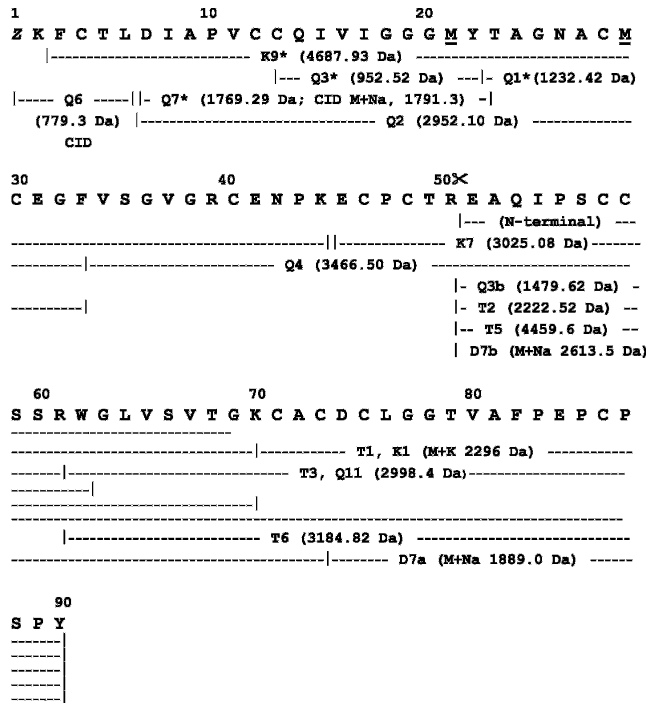

Figure 5. Amino acid sequence of HML. The primary structure of HML was determined by combination of Edman degradation of sets of overlapping peptides obtained by proteolysis of the reduced and carbamidomethylated full-length lectin and its reversed-phase HPLC fragments (isolated as in Fig. 3) with chymotrypsin (Q-), trypsin (T-), endoproteinase Lys-C (K-), and endoproteinase Asp-N (D-), and by CID MS/MS analysis of the N-terminal blocked peptide Q76. Methionine residues at positions 21 and 29 were oxidized in peptides Q1 and K9, respectively. Proteolysis at the Arg50–Glu51 peptide bond, which generates the N-terminal sequence determined in the native two-chain HML lectin and in its 4460 C-terminal fragment (identical to peptide T5), is indicated by scissors. (Z) Pyroglutamic acid. Mass spectrometric sequence determination of Q6 is shown in Figure 6.

formation of a single conformational saccharide recognition surface, awaits the structure elucidation of lectin-carbohydrate complexes.

Materials and methods

Collection of algae and purification of lectins

Specimens of the red algae *H. cervicornis* and *H. musciformis* were collected in the Pacheco beach at the Atlantic coast of the Ceará State of Brazil. The material was cleansed from epiphytes, transported within 1 h of collection to the laboratory, and stored at −20°C until used. The frozen algae were ground to a fine powder in liquid nitrogen. The powder was extracted, while stirring with 5 volumes of 20 mM phosphate buffer (pH 7.0), containing 150 mM NaCl (PBS). Particulate matter was removed by straining through a nylon tissue, followed by centrifugation at 15.000*g* for 20 min at 4°C. For purification of HCA, the supernatant was acidified to pH 1.0 with HCl and left for 5 h at 4°C. This acid treatment effectively removed pigments (phycobilins) that usually interfere with the subsequent chromatographic steps. The precipitated pigments were removed by centrifugation, and the supernatant (crude extract) was adjusted to pH 7.0 with NaOH. Proteins were allowed to precipitate at 25°C for 4 h following the addition of ammonium sulfate to 90% saturation. For purification of HML, an HML-enriched fraction was obtained by ammonium sulfate (70% saturation) precipitation. The precipitated HCA and HML proteins were pelleted by centrifugation, resuspended in a small volume of PBS, dialyzed against 20 mM phosphate buffer (pH 7.0) (PB), and loaded onto a DEAE-Sephacel column equilibrated with the same buffer and eluted at a flow rate of 30 mL/h until the column effluent showed absorbance at 280 nm of < 0.05. The adsorbed proteins were eluted with a linear gradient of 0–2 M NaCl in PB buffer. The elution was monitored at 280 nm, and 3-mL fractions were collected manually and tested for hemagglutinating activity toward rabbit native or trypsinized erythrocytes. Active fractions were pooled, dialyzed extensively against distilled water, freeze-dried, and stored at −30°C until used. The purity of the lectins was assessed by MALDI-TOF mass spectrometry (as below) and N-terminal sequencing (using an Applied Biosystems Precise instrument following the manufacturer's instructions). When necessary, the lectins were further purified by reversed-phase HPLC using a Lichrospher RP100 C18 column (25 × 0.4 cm, 5 μm particle size) eluting at a flow rate of 1.0 mL/min with a mixture of 0.1% (v/v) TFA in water (solvent A) and in acetonitrile (solvent B) using the following chromatographic conditions: first, isocratic (5% B) for 5 min, followed by gradients of 5%–40% B for 10 min, 35%–45% B for 10 min, and 45%–75% B for

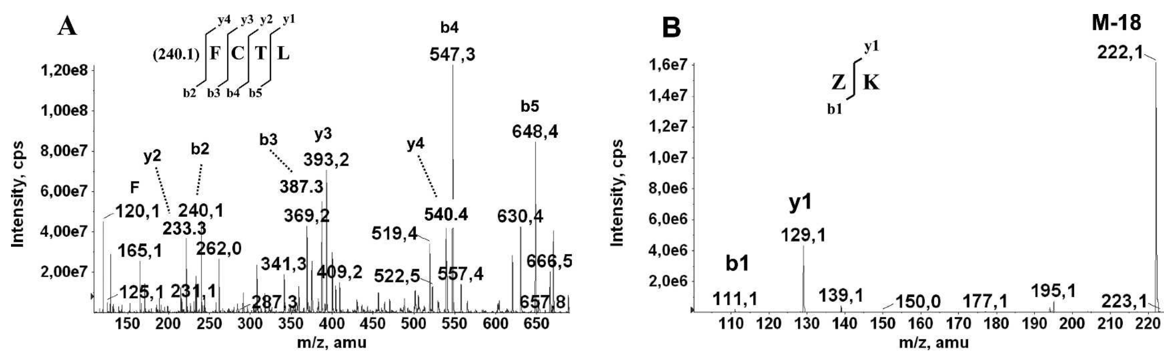

Figure 6. Tandem mass spectrometry. (*A*) Collision-induced fragmentation of the simply charged ion at $m/z = 779.3$ corresponding to the N-terminal blocked peptide Q6 of HML (Fig. 5). Sequence-specific b and y ions used for structure determination are indicated. (*B*) The full $(MS)^3$ spectrum and sequence assignment of the b2 ion at $m/z = 240.1$ produced by MS/MS of the peptide ion Q6 shown in *A*. (F) Immonium ion of phenylalanine at $m/z = 120.1$. (Z) Pyroglutamic acid.

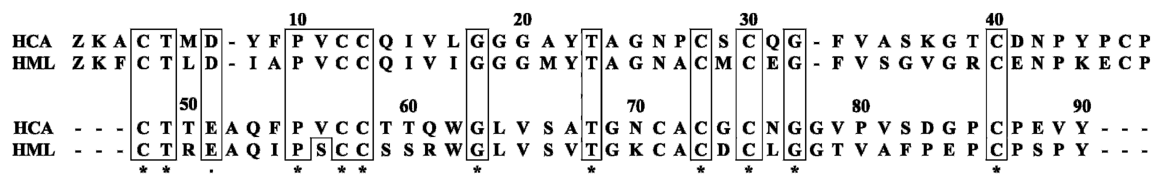

Figure 7. Domain duplication. Amino acid sequence alignment of the N-terminal (residues 1–47) and C-terminal (residues 48–90) tandemly arranged domains of HCA and HML, revealing internal domain duplication. Identical residues are boxed and labeled with an asterisk *below* the sequences.

25 min. Protein elution was simultaneously monitored at 216 and 280 nm, and fractions were collected manually and dried in a vacuum centrifuge (Speed-Vac).

Homogeneity, molecular mass determination, and quantitation of cysteine residues

The purified lectins were visualized in Coomassie blue-stained SDS-(15%) polyacrylamide gels with or without prior reduction with 1% (v/v) 2-mercaptoethanol at 100°C for 2 min. For mass determination and quantitation of sulfhydryl groups and disulfide bonds, the purified proteins (1 μg in 2 μL of 100 mM ammonium bicarbonate [pH 8.3], containing 5 M guanidinium hydrochloride) were incubated with either 10 mM iodoacetamide for 1 h at room temperature, or with 10 mM DTT for 15 in at 65°C, followed by the addition of a fivefold molar excess of iodoacetamide over reducing agent and incubation for 1 h at room temperature. The reaction mixtures were freed from reagents using a C18 Zip-Tip pipette (Millipore) after activation with 70% acetonitrile and equilibration in 0.1% trifluoroacetic acid (TFA). Following protein adsorption and washing with 0.1% TFA, the proteins were eluted onto the MALDI-TOF plate with 1 μL of 70% acetonitrile and 0.1% TFA and subjected to mass spectrometric analysis. The molecular masses of the native and the reduced and carbamidomethylated lectins were determined by MALDI-TOF mass spectrometry using an Applied Biosystems Voyager DE-PRO instrument operating at 25 kV accelerating voltage in the linear mode, and using 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) saturated in 70% acetonitrile and 0.1% TFA as the matrix. The mass calibration standard consisted of a mixture of the following proteins, whose isotope-averaged molecular masses in daltons are given in between brackets: bovine insulin (5734.5), *Escherichia coli* thioredoxin (11,674.5), and horse apomyoglobin (16,952.6).

The number of free cysteine residues (N_{SH}) was determined using the equation:

$$N_{SH} = (M_{IA} - M_{NAT})/57.05 \quad (1)$$

where M_{IA} is the mass of the denatured but nonreduced protein incubated in the presence of iodoacetamide, M_{NAT} is the mass of the native protein, and 57.05 is the mass increment due to the carbamidomethylation of one thiol group.

The number of total cysteine residues (N_{Cys}) was derived using:

$$N_{Cys} = [(M_{CM} - M_{IA})/58.05] + N_{SH} \quad (2)$$

where M_{CM} is the mass of the reduced and carbamidomethylated protein and 58.05 is the mass increment due to the carbamidomethylation of a cysteine residue, which prior to reduction was involved in the formation of a disulfide bond.

Finally, the number of disulfide bonds ($N_{S\text{–}S}$) was calculated from:

$$N_{S\text{-}S} = (N_{cys} - N_{SH})/2 \quad (3)$$

All mass values in equations I-III are in daltons.

Amino acid sequence determination

The primary structures of HCA and HML were established by N-terminal sequence analysis of reversed-phase HPLC-purified reduced and carbamidomethylated fragments, and of sets of overlapping peptides obtained by proteolytic digestions. To this end, 100 μg of each purified protein was dissolved in 100 mM NH_4HCO_3 (pH 8.3) and subjected to proteolysis by trypsin, chymotrypsin, endoproteinase Lys-C, and endoproteinase Asp-N (at an enzyme-to-protein ratio of 1:100, w/w) overnight at 37°C. Peptides were fractionated by reverse-phase HPLC on a Vydac C_{18} (4.6 × 250 mm) column equilibrated in 0.1% (v/v) trifluoroacetic acid (TFA) in water. Elution was performed at a flow rate of 0.8 mL/min with a linear gradient of 0%–80% acetonitrile in 0.1% TFA for 100 min. Peptides were characterized by N-terminal sequence analysis (using an Applied Biosystems Precise instrument following the manufacturer's instructions) and matrix-assisted laser-desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) using an Applied Biosystems Voyager DE-Pro spectrometer and α-cyano-4-hydroxycinnamic acid (saturated in 70% acetonitrile and 0.1% TFA as matrix) as the matrix. A tryptic peptide mixture from the *Cratylia floribunda* seed lectin (SwissProt accession code P81517) prepared and previously characterized in our laboratory was used as the mass calibration standard (mass range 450–3300 Da).

The amino acid sequences of the N-terminal-blocked chymotryptic peptides Q7 (m/z = 2537.06; HCA) and Q6 (m/z = 779.31; HML) were determined by collision-induced dissociation (CID) tandem mass spectrometry, MS/MS, and $(MS)^3$, using a linear ion trap (Qtrap; Applied Biosystems) mass spectrometer (Hager and Le Blanc 2003) equipped with a nanoelectrospray source (Protana). The CID spectra were interpreted manually.

Amino acid sequence and disulfide-bonding similarity searches

Amino acid sequence similarity searches were carried out against a nonredundant protein databank using the program PSI-BLAST (Altschul et al. 1997) available at http://www.ncbi.nlm.nih.gov/

BLAST. Possible cysteine-pairing patterns and disulfide bond connectivity similarities were searched using the CysView program (Lenffer et al. 2004; available at http://research.i2r.a-star.edu.sg/CysView/) against the nonredundant UniProt database downloaded from the Expasy FTP server (ftp://au.expasy.org/databases/uniprot/current_release/knowledgebase/complete/), and against the Disulphide Database (DSDBASE) (Vinayagam et al. 2004; at http://www.ncbs.res.in/~faculty/mini/dsdbase/dsdbase.html).

Hemagglutination and hemagglutination-inhibition tests

D-Glucose, D-mannose, D-galactose, methyl-α-D-galactopyranoside, L-fucose, *N*-acetyl-D-glucosamine, *N*-acetyl-D-galactosamine, lactose, lactulose, carrageenan, fucoidan, porcine stomach mucin, bovine submaxillary mucin, bovine fetuin and asialofetuin, hen ovalbumin and ovomucoid, porcine thyroglobulin, and yeast mannan were purchased from Sigma Aldrich Corp. (USA). Human lactotransferrin and serotransferrin and bovine lactotransferrin were gifts from Dr. G. Spik (USTL). Ovine submaxillary mucin was isolated according to the method of Hill et al. (1977). Asialoglycoproteins were prepared by treatment with 0.1 N trifluoracetic acid for 1 h at 80° C, dialysis against distilled water, and lyophilization.

Agglutination of either 3% native or trypsin-treated rabbit red blood cell suspension in PBS by HML and HCA and inhibition of this agglutination activity by various simple sugars or glycoconjugates were carried out in U-bottom microtiter plates (Thermo Labsystems) by a twofold serial dilution technique. In each tube, 50 μL of a twofold serial dilution of simple sugars or glycoconjugates in PBS was added to an equal volume of lectin solution, which had been carefully diluted to contain four minimum agglutination doses. After 1 h at room temperature, 50 μL of the erythrocyte suspension was added. The mixture was left for 1 h at room temperature and then examined for agglutination. Results were expressed as the minimum concentration of simple sugars (millimolar) or glycoproteins (micrograms per milliliter) required to completely inhibit four hemagglutinating units.

Acknowledgments

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Financiadora de Estudos e Projetos (FINEP), Fundação Cearense de Amparo à Pesquisa (FUNCAP), and grants CAPES/COFECUB 336/01, and BFU2004-01432/BMC from the Ministerio de Educación y Ciencia, Madrid (Spain). C.S.N. is the recipient of a fellowship from the Coordenação Aperfeiçoamento de Pessoal de Nivel Superior (CAPES). A.H.S. and B.S.C. are senior investigators of CNPq.

References

Ainouz, I.L. and Sampaio, A.H. 1991. Screening of Brazilian marine algae for hemagglutinins. *Bot. Mar.* **34:** 211–214.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.

Ambrosio, A.L., Sanz, L., Sánchez, E.I., Wolfenstein-Todel, C., and Calvete, J.J. 2003. Isolation of two novel mannan- and L-fucose-binding lectins from the green alga *Enteromorpha prolifera*: Biochemical characterization of EPL-2. *Arch. Biochem. Biophys.* **415:** 245–250.

Boyd, W.C., Almodovar, L.R., and Boyd, L.G. 1966. Agglutinins in marine algae for human erythrocytes. *Transfusion* **6:** 82–83.

Calvete, J.J., Costa, F.H.F., Saker-Sampaio, S., Murciano, M.P., Nagano, C.S., Cavada, B.S., Grangeiro, T.B., Ramos, M.V., Bloch Jr., C., Silveira, S.B., et al. 2000. The amino acid sequence of the agglutinin isolated from the red marine alga *Bryothamnion triquetrum* defines a novel lectin structure. *Cell. Mol. Life. Sci.* **57:** 343–350.

Chai, W., Hounsell, E.F., Cashmore, G.C., Rosankiewicz, J.R., Feeney, J., and Lawson, A.M. 1992. Characterization by mass spectrometry and ^{1}H-NMR of novel hexasaccharides among the acidic *O*-linked carbohydrate chains of bovine submaxillary mucin. *Eur. J. Biochem.* **207:** 973–980.

Codeville, B., Strecker, G., Wieruszeski, J.M., Vliegenthart, J.F.G., Van Halbeek, H., Peter-Katalinic, J., Egge, H., and Spik, G. 1992. Heterogeneity of bovine lactotransferrin glycans. Characterization of α-D-Galp(1 → 3)-β-D-Gal- and α-NeuAc-(2 → 6)-β-D-GalpNAc-(1 → 4)-β-D-GlcNAc-substituted N-linked glycans. *Carbohydr. Res.* **236:** 145–164.

De Waard, P., Koorevaar, A., Kamerling, J.P., and Vliegenthart, J.F.G. 1991. Structure determination by ^{1}H NMR spectroscopy of (sulfated) sialylated N-linked carbohydrate chains released from porcine thyroglobulin by peptide-N^{4}-(N-acetyl-β-glucosaminyl) asparagine amidase-F. *J. Biol. Chem.* **266:** 4237–4243.

Elgavish, S. and Shaanan, B. 1997. Lectin-carbohydrate interactions: Different folds, common recognition principles. *Trends Biochem. Sci.* **22:** 462–467.

Gabius, H.-J. and Gabius, S., eds. 1997. *Glycoscience. Status and perspectives*. Chapman and Hall, Weinheim, Germany.

Hager, J.W. and Le Blanc, Y.J.C. 2003. Product ion scanning using a Q-q-Q linear ion trap (Q TRAP) mass spectrometer. *Rapid Commun. Mass Spectrom.* **17:** 1056–1064.

Hill Jr., H.D., Reynolds, J.A., and Hill, R.L. 1977. Purification, composition, molecular weight, and subunit structure of ovine submaxillary mucin. *J. Biol. Chem.* **252:** 3791–3798.

Hori, K., Matsubara, K., and Miyazawa, K. 2000. Primary structures of two hemagglutinins from the marine red alga, *Hypnea japonica*. *Biochim. Biophys. Acta* **1474:** 226–236.

Karlsson, N.G., Nordman, H., Karlsson, H., Carlstedt, I., and Hansson, G.C. 1997. Glycosylation differences between pig gastric mucin populations: A comparative study of the neutral oligosaccharides using mass spectrometry. *Biochem. J.* **326:** 911–917.

Kawakubo, A., Makino, H., Ohnishi, J.-I., Hirohara, H., and Hori, K. 1999. Occurrence of highly yielded lectins homologous within the genus *Eucheuma*. *J. Appl. Phycol.* **11:** 149–156.

Lenffer, J., Lai, P., El Mejaber, W., Khan, A.M., Koh, J.L., Tan, P.T., Seah, S.H., and Brusic, V. 2004. CysView: Protein classification based on cysteine pairing patterns. *Nucleic Acids Res.* **32 (Web Server issue):** W350–W355.

Loris, R. 2002. Principles of structures of animal and plant lectins. *Biochim. Biophys. Acta* **1572:** 198–208.

Nagano, C.S., Moreno, F.B.M.B., Bloch Jr., C., Prates, M.V., Calvete, J.J., Saker-Sampaio, S., Farias, W.R.L., Tavares, T.D., Nascimento, K.S., Grangeiro, T.B., et al. 2002. Purification and characterization of a new lectin from the red marine alga *Hypnea musciformis*. *Prot. Pept. Lett.* **9:** 159–165.

Rademacher, T.W., Homans, S.W., Parekh, R.B., and Dwek, R.A. 1986. Immunoglobulin G as a glycoprotein. *Biochem. Soc. Symp.* **51:** 131–148.

Rogers, D.J. and Hori, K. 1993. Marine algal lectins: New developments. *Hydrobiologia* **260/261:** 589–593.

Sampaio, A.H., Rogers, D.J., and Barwell, C.J. 1998. A galactose-specific lectin from the red marine alga *Ptilota filicina*. *Phytochemistry* **48:** 765–769.

Sampaio, A.H., Rogers, D.J., Barwell, C.J., Saker-Sampaio, S., Costa, F.H.F., and Ramos, M.V. 1999. A new isolation procedure and further characterisation of the lectin from the red marine alga *Ptilota serrata*. *J. Appl. Phycol.* **10:** 539–546.

Sampaio, A.H., Rogers, D.J., Barwell, C.J., Saker-Sampaio, S., Nascimento, K.S., Nagano, C.S., and Farias, W.R.L. 2002. New affinity procedure for the isolation and further characterization of the blood group B specific lectin from the red marine alga *Ptilota plumosa*. *J. Appl. Phycol.* **14:** 489–495.

Savage, A.V., Donohue, J.J., Koeleman, C.A., and Van Den Eijnden, D.H. 1990. Structural characterization of sialylated tetrasaccharides and pentasaccharides with blood group H and Le(x) activity isolated from bovine submaxillary mucin. *Eur. J. Biochem.* **193:** 837–843.

Savage, A.V., D'Arcy, S.M.T., and Donoghue, C.M. 1991. Structural characterization of neutral oligosaccharides with blood group A and H activity isolated from bovine submaxillary mucin. *Biochem. J.* **279:** 95–103.

Spik, G., Bayard, B., Fournet, B., Strecker, G., Bouquelet, S., and Montreuil, J. 1975. Studies on glycoconjugates. LXIV. Complete structure of two carbohydrate units of human serotransferrin. *FEBS Lett.* **50:** 296–299.

Spik, G., Strecker, G., Fournet, B., Bouquelet, S., Montreuil, J., Dorland, L., Van Halbeek, H., and Vliegenthart, J.F.G. 1982. Primary structure of the glycans from human lactotransferrin. *Eur. J. Biochem.* **121:** 413–419.

Van Damme, E.J.M., Peumans, W.J., Barre, A., and Rougé, P. 1998. Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Crit. Rev. Plant Sci.* **17:** 575–692.

Van Halbeek, H., Dorland, L., Vliegenthart, J.F.G., Kochetkov, K., Arbatsky, N.P., and Derevitskaya, V.A. 1982. Characterization of the primary structure and the microheterogeneity of the carbohydrate chains of porcine blood-group H substance by 500-MHz ^{1}H-NMR spectroscopy. *Eur. J. Biochem.* **127:** 21–29.

Vinayagam, A., Pugalenthi, G., Rajesh, R., and Sowdhamini, R. 2004. DSDBASE: A consortium of native and modelled disulphide bonds in proteins. *Nucleic Acids Res.* **32:** D200–D202.

Wang, S., Zhong, F.-D., Zhang, Y.-J., Wu, Z.-J., Lin, Q.-Y., and Xie, L.-H. 2004. Molecular characterization of a new lectin from the marine alga *Ulva pertusa*. *Acta Biochim. Biophys. Sinica* **36:** 111–117.

Zenteno, E., Vasquez, R., Cordoba, F., Wieruszeski, J.M., Montreuil, J., and Debray, H. 1995. Specificity of the isolectins from the plant cactus *Machaerocereus eruca* for oligosaccharides from porcine stomach mucin. *Glycoconjugate J.* **12:** 699–706.