CMLS, Cell. Mol. Life Sci. 57 (2000) 343–350
1420-682X/00/020343-08 $ 1.50 + 0.20/0

CMLS Cellular and Molecular Life Sciences

Research Article

The amino acid sequence of the agglutinin isolated from the red marine alga *Bryothamnion triquetrum* defines a novel lectin structure

J. J. Calvete[1,*], F. H. F. Costa[2], S. Saker-Sampaio[2], M. P. M. Murciano[1], C. S. Nagano[2], B. S. Cavada[2], T. B. Grangeiro[2], M. V. Ramos[2], C. Bloch Jr[3], S. B. Silveira[2], B. T. Freitas[2] and A. H. Sampaio[2]

[1]Instituto de Biomedicina de Valencia, Jaime Roig 11, E-46010 Valencia (Spain), Fax + 34 96 369 0800, e-mail: jcalvete@ibv.csic.es
[2]Departamento de Engenharia de Pesca & BioMol-Lab, Universidade Federal do Ceará, C.P. 6033, CEP 60451-970, Fortaleza-Ceará (Brazil)
[3]EMBRAPA/CENARGEN, Brasilia-DF (Brazil)

Received 4 December 1999; accepted 13 December 1999

Abstract. The primary structure of a lectin isolated from the red alga *Bryothamnion triquetrum* was established by combination of Edman degradation of sets of overlapping peptides and mass spectrometry. It contains 91 amino acids and two disulphide bonds. The primary structure of the *B. triquetrum* lectin does not show amino acid sequence similarity with known plant and animal lectin structures. Hence, this protein may be the paradigm of a novel lectin family.

Key words. Red marine alga; *Bryothamnion triquetrum*; lectin; amino acid sequence; mass spectrometry.

Lectins are a structurally very diverse class of carbohydrate-binding proteins widely distributed in nature and found in animals, plants and micro-organisms. Lectins interact with specific glycan structures linked to soluble and membrane-bound glycoconjugates, and these protein-carbohydrate interactions play key roles in biological processes such as cell communication, pathogen infection and host defence, fertilization, development, cell differentiation and cancer metastasis [1–5]. Though in many instances their exact biological roles remain elusive, lectins from animals and terrestrial plants have been extensively exploited as biochemical tools in biotechnology and biomedical research [6–8].

Lectins have traditionally represented a paradigm for studying protein-carbohydrate interactions. The structural basis of selective carbohydrate recognition by different legume, bacterial, viral and animal lectins has been assessed mainly by X-ray crystallography [5, 9–13; consult also the 3D lectin database at: http://www.cermav.cnrs.fr/databank/lectine/]. In contrast, marine algal lectins have been isolated and characterized from only a few species and at a much slower pace since the first report, more than 30 years ago, of the haemagglutinating activity in these organisms [14]. This paucity is mainly due to difficulties in obtaining sufficient material for study [15]. Hence, little biochemical and structural information is available on this class of lectins.

Most algal lectins are monomeric, low-molecular-mass proteins exhibiting binding specificity for complex oligosaccharides or glycoproteins, and they do not appear to require divalent cations for structural integrity

* Corresponding author.

or biological activity [16–18]. Moreover, no primary structure of algal lectins has been reported, and thus their classification based on evolutionary relationships to known lectin families remains enigmatic. We have reported the isolation of low-molecular-mass (< 10 kDa) monomeric lectins from the red marine algae *Bryothamnion triquetrum* (BTL) and *B. seaforthii* (BSL) [19]. BTL agglutinates trypsin-treated erythrocytes from rabbit, chicken, goat, pig and human ABO erythrocytes, while BSL haemagglutinates trypsinized red cells from rabbit, chicken and cow [20]. The haemagglutinating activities of BTL and BSL are not dependent on divalent cations and are not inhibited by simple sugars [19]. Here we report the complete primary structure of BTL. It does not display similarity with known lectin structures, indicating that it may belong to a novel protein family.

Materials and methods

Purification BTL. Specimens of the red marine alga *B. triquetrum* were collected on the Atlantic coast of Brazil (Pacheco Beach, Ceará) and kept at -20 °C in sealed plastic bags. For purification of BTL, the algae were thawed, rinsed with distilled water, cleaned of epiphytes, ground to a fine powder under liquid nitrogen, extracted with three volumes of 20 mM sodium phosphate, pH 7.0, containing 150 mM NaCl for 4 h under stirring, filtered through nylon tissue and centrifuged at $7000 \times g$ for 30 min at 4 °C. The supernatant was acidified and left at 4 °C for 4 h. Precipitated pigments were removed by centrifugation and the supernatant was adjusted to pH 7.0, and subjected to fractionated precipitation with 60% ammonium sulphate saturation. Precipitated proteins were pelleted by centrifugation, dissolved in a small volume of 20 mM phosphate buffer pH 7.0, and applied to a DEAE-cellulose column equilibrated in the same buffer. The flow-through fraction containing the haemagglutinating activity was rechromatographed on the same column, dialysed against water, and lyophilized. The degree of protein purity was assessed by SDS (12.5%)-polyacrylamide gel electrophoresis, reverse-phase high-performance liquid chromatography (HPLC) on a Lichrospher RP100 C18 column (25×0.4 cm, 5-μm particle size) eluted at 1 ml/min with a mixture of 0.1% (v/v) TFA in water (solvent A) and acetonitrile (solvent B), N-terminal sequencing, and mass spectrometry (see below).

Analytical techniques. N-terminal sequence analyses were performed with either an Applied Biosystems 477A sequencer or a Beckman Porton LF-3000 instrument, following the manufacturer's instructions.

Amino acid and amino sugar analyses were carried out with a Beckman Sytem Gold amino acid analyser after sample hydrolysis in sealed, evacuated ampoules at 110 °C with 6 M HCl for 24 h and with 4 M HCl for 4 h, respectively.

Molecular mass was determined by MALDI-TOF mass spectrometry using a PE Biosystems Voyager DE-STR instrument operating at a 20-kV accelerating voltage in the reflector mode. The sample matrix was 3,5-dimethoxy-4-hydroxycinnamic acid. Horse skeletal apomyoglobin and *Escherichia coli* oxidized thioredoxin were used as internal mass standards.

Determination of sulphydryl groups and disulphide bonds. For quantitation of free cysteine residues and disulphide bonds in BTL, the protein (1 mg/ml in 150 mM Tris/HCl, pH 8.6, 1 mM EDTA, 6 M guanidine hydrochloride) was incubated with either 10 mM iodoacetamide for 1 h at room temperature, or with 1% 2-mercaptoethanol for 2 min at 100 °C, followed by addition of a fivefold molar excess of 4-vinylpyridine over reducing agent and incubation for 1 h at room temperature. Samples were dialysed against deionized (MilliQ) water and lyophilized, and were subjected to MALDI-TOF-MS and amino acid analysis.

Proteolytic cleavages and characterization of fragments. Samples of BTL (2 mg/ml in 100 mM NH_4HCO_3, pH 8.6, containing 1 M guanidine hydrochloride and 5 mM $CaCl_2$) were digested with α-chymotrypsin, thermolysin, elastase and endoproteinase Lys-C (1:100 w/w enzyme:substrate ratio) for 18 h at 37 °C. Proteolytic peptides were isolated by reverse-phase HPLC on a Lichrospher RP100 C18 column (25×0.4 cm, 5-μm particle size) eluting at 1 ml/min with a mixture of 0.1% (v/v) TFA in water (solvent A) and acetonitrile (solvent B) employing the following chromatographic conditions: first, isocratic (10% B) for 5 min, followed by gradients of 10–40% B for 90 min, and 40–70% B for 30 min. Peptide elution was monitored at 220 nm and chromatographic fractions were collected manually and characterized by amino acid analysis, N-terminal sequencing and mass spectrometry.

Similarity searches. Amino acid sequence similarity searches were carried out against a non-redundant protein databank using the program BLAST [21] (available at http://www.ncbi.nlm.nih.gov/BLAST).

Results and discussion

Amino acid sequence of BTL. Purified BTL migrated in SDS-polyacrylamide gels as a major broad electrophoretic band with an apparent molecular mass of 6–8 kDa and a minor band at around 16 kDa (fig. 1, insert). The latter band disappeared upon treatment with 2-mercaptoethanol, suggesting that it could represent a dimer of the low-molecular-mass band. Indeed, N-terminal amino acid sequence analysis of both the unfractionated sample and the electroblotted 16 and 6-

to 8-kDa bands showed the single sequence: AD-PIXG(G + S)SGYSXTTPAILTPKSPGSFPSGYSKVI-VT. Determination of the molecular mass of native BTL by MALDI-TOF mass spectrometry showed major ions at m/z 8980.5, 8964.7 and 8949.4, and smaller amounts of higher-molecular-mass species at m/z 17958, 26938, 35923, etc. (fig. 1A). Since BTL yielded a single N-terminal sequence, this result further suggested the existence of lectin isoforms and the presence of multimeric forms.

Sequence heterogeneity was found in position 7, and residues at positions 5 and 12 could not be identified. Galactosamine and glucosamine were not detected by amino sugar analysis (table 1), indicating that BTL may not be a glycoprotein. However, 3.4 moles % cysteine was quantified by amino acid analysis. This indicated that the unidentified residues could be cysteine residues. This point was confirmed by N-terminal sequence analysis of reduced and ethylpyridylated BTL. Amino acid analysis of non-reduced lectin, which had been incubated under denaturing conditions with vinylpyridine, failed to show the presence of ethylpyridylcysteine, indicating that BTL does not possess free sulphydryl groups. On the other hand, cysteine was quantitatively recovered as the ethylpyridyl derivative in the amino acid analysis of reduced and alkylated lectin (table 1). Together, these results indicated the existence of four cysteine residues per lectin monomer and that these residues are engaged in the formation of two intramolecular disulphide bridges. This was confirmed by mass spectrometry. Thus, the molecular masses of the major BTL ions did not change upon incubation of the lectin with vinylpyridine under denaturing but non-reducing conditions (fig. 1B). However, mass spectrometric analysis of ethylpyridylated BTL showed three major ions at m/z 9379.5, 9406, and 9420 (fig. 1C). The mass increment of 436 ± 6 Da of each native BTL

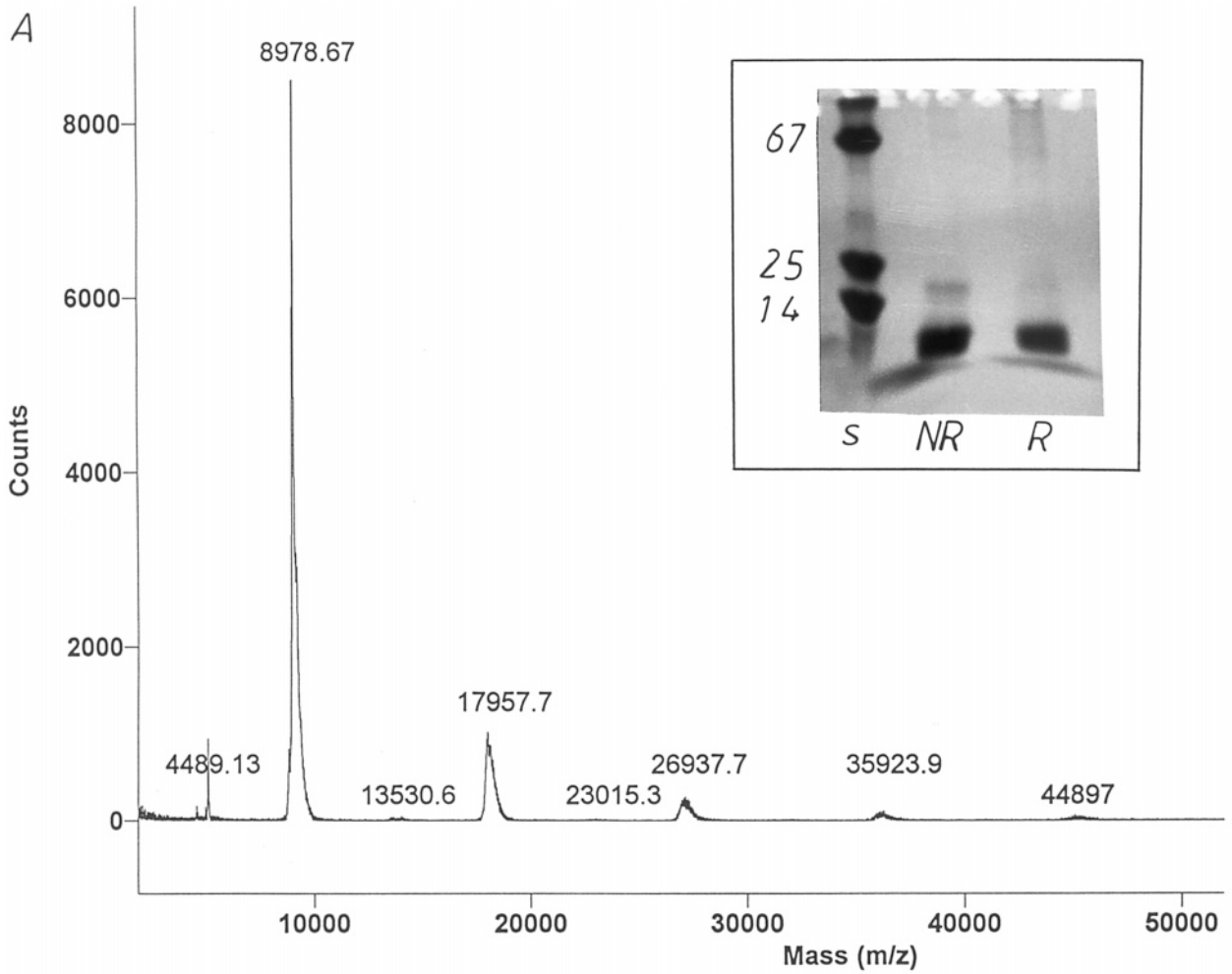

Fig. 1(*A*).

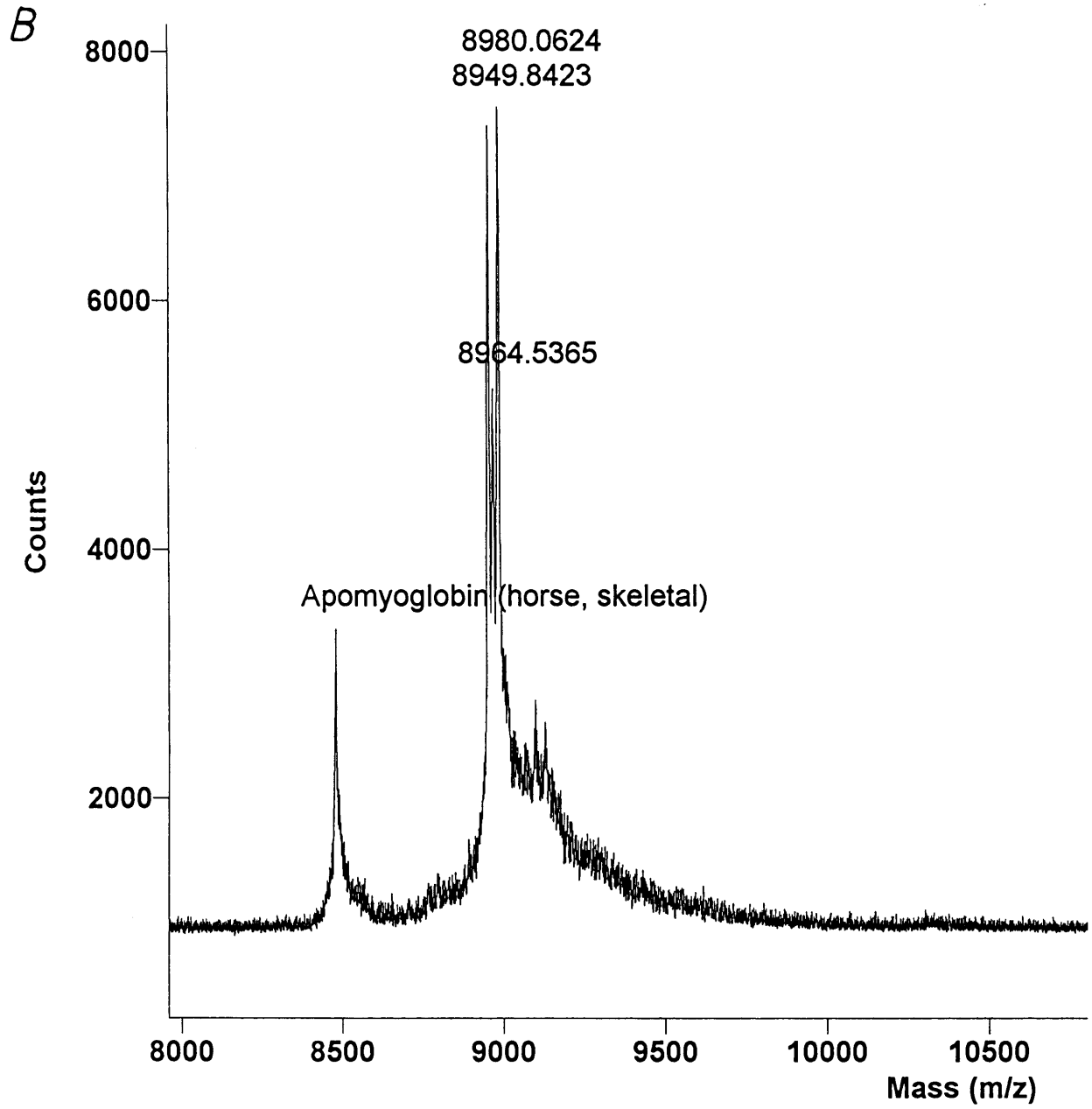

Fig. 1(*B*).

species may correspond to the alkylation of four cysteine residues with ethylpyridyl groups (calculated mass increment of 425.2 Da), indicating the presence of two intramolecular disulphide bonds per BTL molecule.

The amino acid sequence of BTL was established after N-terminal sequence analysis of sets of overlapping peptides obtained by proteolytic digestions of the ethylpyridylated protein (fig. 2). Heterogeneity was found at positions 7 (S/G), 39 (A/G), 53 (A/K), 55 (A/Y), 65 (F/I) and 82 (S/G), confirming that the heterogeneity found by mass spectrometry is due to the presence of isolectins. The calculated isotope-averaged molecular mass of the 91-amino acid residue sequence shown in figure 2, with G^{39}, K^{53}, Y^{55}, F^{65} and a glycine residue at position 7 or 82, and taking into account that the four cysteine residues form two disulphide bonds, is 8981 Da, which is in very good agreement with the experimentally determined mass. The calculated molec-

ular mass of the amino acid sequence shown in figure 2 with A^{39} and glycine residues at both positions 7 and 82 is 8967 Da, and may thus correspond to one of the other major BTL isoforms detected by MALDI-TOF-MS (fig. 1A). On the other hand, the third major BTL isoform of 8949 Da (fig. 1A) may contain the same amino acid sequence as the 8981-Da species except for the F/I substitution at position 65 (calculated isotope-averaged mass of 8947 Da). Amino acid analysis of the BT isolectin mixture (table 1) is in very good agreement with the composition calculated from the amino acid sequence.

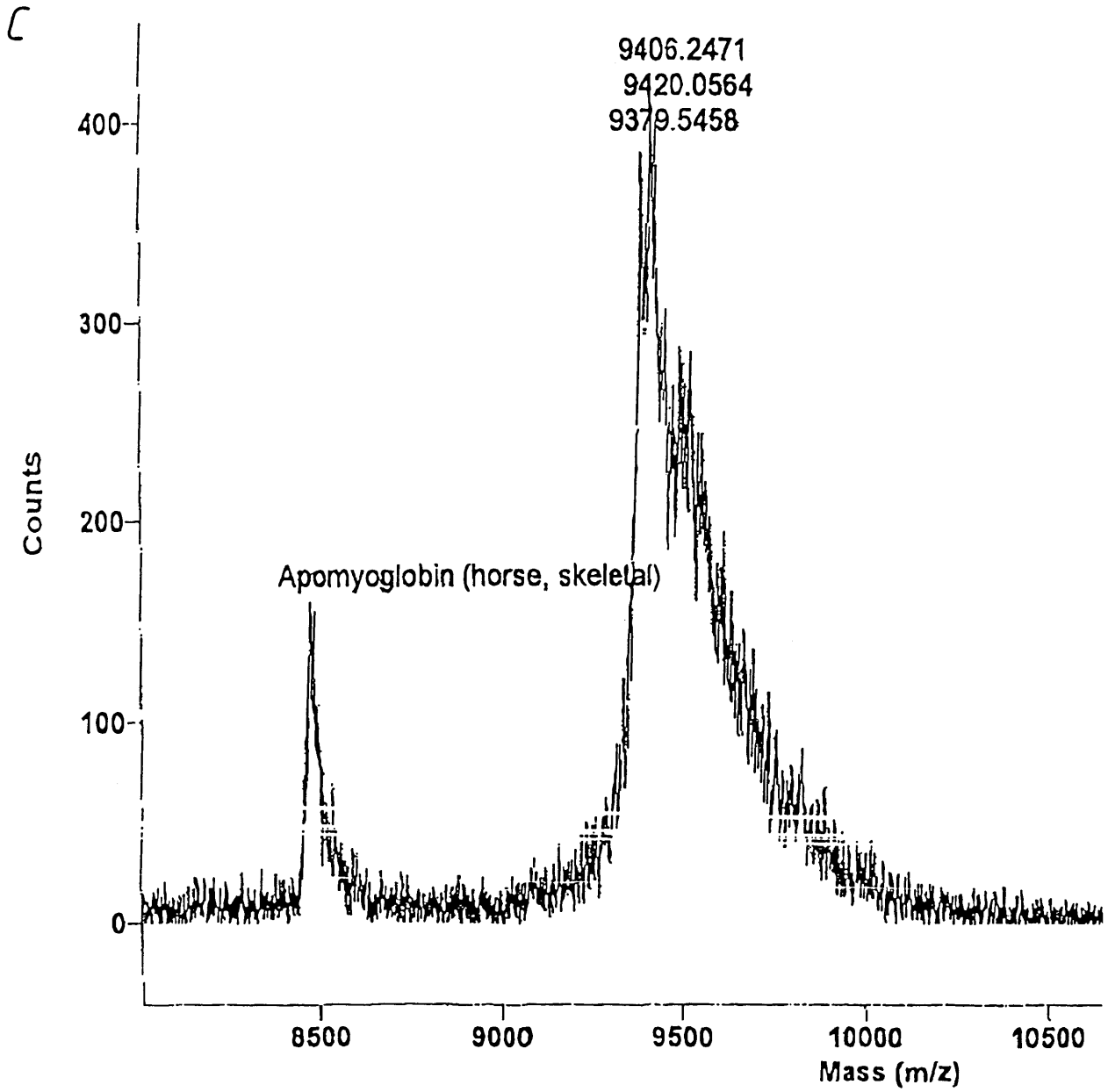

Figure 1. (*A*) Matrix-assisted laser-desorption ionization time-of-flight (MALDI-TOF) mass spectrometric analysis of purified BTL showing a major quasimolecular ion at m/z 8978.7 and small amounts of its multimers at m/z 17957.7 (dimer), 26937.7 (trimer), 35923.9 (tetramer). Insert, lane s, molecular mass standards, from top to bottom, bovine serum albumin (68 kDa), chymotrypsinogen (25 kDa), hen egg lysozyme (14 kDa). Lane NR, SDS (12.5%)-polyacrylamide gel electrophoresis of non-reduced BTL. Lane R, SDS (12.5%)-polyacrylamide gel electrophoresis of reduced BTL. (*B*) Mass spectrum of BTL after incubation with iodoacetamide under denaturing, non-reducing conditions. (*C*) Mass spectrum of BTL after reduction and pyridylethylation.

Table 1. Comparison of the amino acid composition of the *Bryothamnion triquetrum* lectin determined by amino acid analysis and calculated from the amino acid sequence.

Amino acid sequence	Residues/mol	Residues from
Asx	3.5	3
Thr	5.3	5
Ser*	16.3	15–16
Glx	2.4	2
Gly*	18.0	16–19
Pro	6.3	6
Ala*	7.0	6–9
Cys	3.6	4
Val	7.8	9
Met	–	–
Ile*	3.8	4–5
Leu	1.3	1
Tyr*	6.3	6–7
Phe*	2.5	2–3
His	0.7	1
Lys*	5.3	5–6
Arg	0.9	1
GlcN	–	–
GalN	–	–

Cysteine was determined as its ethylpyridyl derivative. Asx, aspartic acid plus asparagine; Glx, glutamic acid plus glutamine; GlcN, glucosamine; GaLN, galactosamine. For calculation of residues/mol of the *B. triquetrum* isolectin mixture, an averaged molecular mass of 8965 Da was assumed. The calculated number of residues labelled with an asterisk varies due to amino acid sequence heterogeneity.

Location of disulphide bridges. For location of disulphide bonds, native BTL was degraded with endoproteinase Lys-C and the resulting peptides were isolated by reversed-phase HPLC and characterized by N-terminal sequencing and MALDI-TOF-MS. Sequence analysis showed that fragment K4 contained a major N-terminal amino acid sequence (^{1}ADPIXGSSGYSXTTPAILTPK^{21}) although the minor sequence ^{11}SXTTPAILTPK^{21} was clearly identified. Mass spectrometric analysis of this fragment showed quasimolecular ions at m/z 2081 and 2099. The 2081-Da ion may correspond to the 21-residue polypeptide chain with an intrachain disulphide bond (calculated isotope-averaged mass of 2081.4 Da), and the 2099 Da ion was assigned to peptides 1–10 and 11–21 linked by an interchain disulphide bridge. Another fragment, K9, had the amino acid sequence ^{1}ADPIXGGSGYSXT and a molecular mass of 5297 Da. These data clearly identified K9 as the polypeptide 1–52 with serine and alanine at positions 7 and 39, respectively, and an intramolecular disulphide bond (calculated isotope-averaged molecular mass of 5296 Da). On the other hand, fragment K5 contained a single amino acid sequence (ASEGGXASFG), which was assigned to a polypeptide starting at position 57 (fig. 2). Its molecular mass of 3232 Da indicated that this fragment corresponded to the polypeptide stretch 57–91 including an intrachain disulphide bond. These results indicated the existence of disulphide linkages between neighbouring cysteine residues in the pattern 1–2 and 3–4.

Native BTL has been described as a monomeric protein of molecular mass lower that 10 kDa [19]. However, the small amounts of multimer of the lectin detected by mass spectrometry indicated that the protein is also capable of forming n-mer aggregates, most probably by disulphide interchange between subunits. In line with this conclusion, a minor proteolytic peptide corresponding to amino acid sequences 57–68 (ASEGGCASFGSY, $M + H^+ = 1135.5$ Da) was characterized. This fragment includes cysteine residue 62, suggesting that the disulphide bond Cys62–Cys90 is labile.

BTL defines a novel protein family. When the amino acid sequence of BTL was compared with all other protein sequences deposited in public databases, no significant similarity was found. This strongly suggests that BTL is not evolutionarily related to any other known lectin structure, and may, therefore, be a member of a new lectin family. Recently, Kawabuko et al. [22] reported the isolation and partial structural characterization of isolectins from three species of the red alga *Eucheuma* (*E. serra*, *E. amakusaensis* and *E cottonii*). The N-terminal amino acid sequences of these isolectins (molecular mass of 29 kDa) were almost identical [GRYTV(Q/K)NQWGGSSAPWNDAG]. We have obtained the same N-terminal sequence for a 30-kDa lectin isolated from the Brazilian red marine alga *Meristiela echinocarpa* [unpublished results]. These sequences do not show any sequence similarity with the *B. triquetrum* lectin reported here. Moreover, we have determined the amino acid sequences of proteolytic peptides derived from another red marine alga (*Ptilota serrata*) lectin [unpublished data], an 18-kDa polypeptide [23]. The peptide sequences cover about 70% of its primary structure and do not show discernible amino acid sequence similarity with either the *Euchema*/*Meristiela* or with the *Bryothamnion* lectins. This strongly suggests the existence of structurally different lectins in red marine algae, which in turn do not resemble known lectins of higher plants or animals. Whether marine algae and terrestrial plants share evolutionarily related lectins, or their carbohydrate-binding activities evolved by convergent evolution awaits the structural elucidation of more lectins from different algal genera.

Acknowledgements. This work was financed by the following: Brazilian Institutions Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)/COFECUB, BioTools Ecological Foundation, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Financiadora de Estudos e Projetos (FINEP), Programa de Apoio ao Desnvolvimento Científico e Tecnológico (PADCT), Fundação Cearense de Amparo á Pesquisa (FUNCAP-CE), and Banco do Nordeste, and by grants from the International Foundation for Science, and PB97-1237

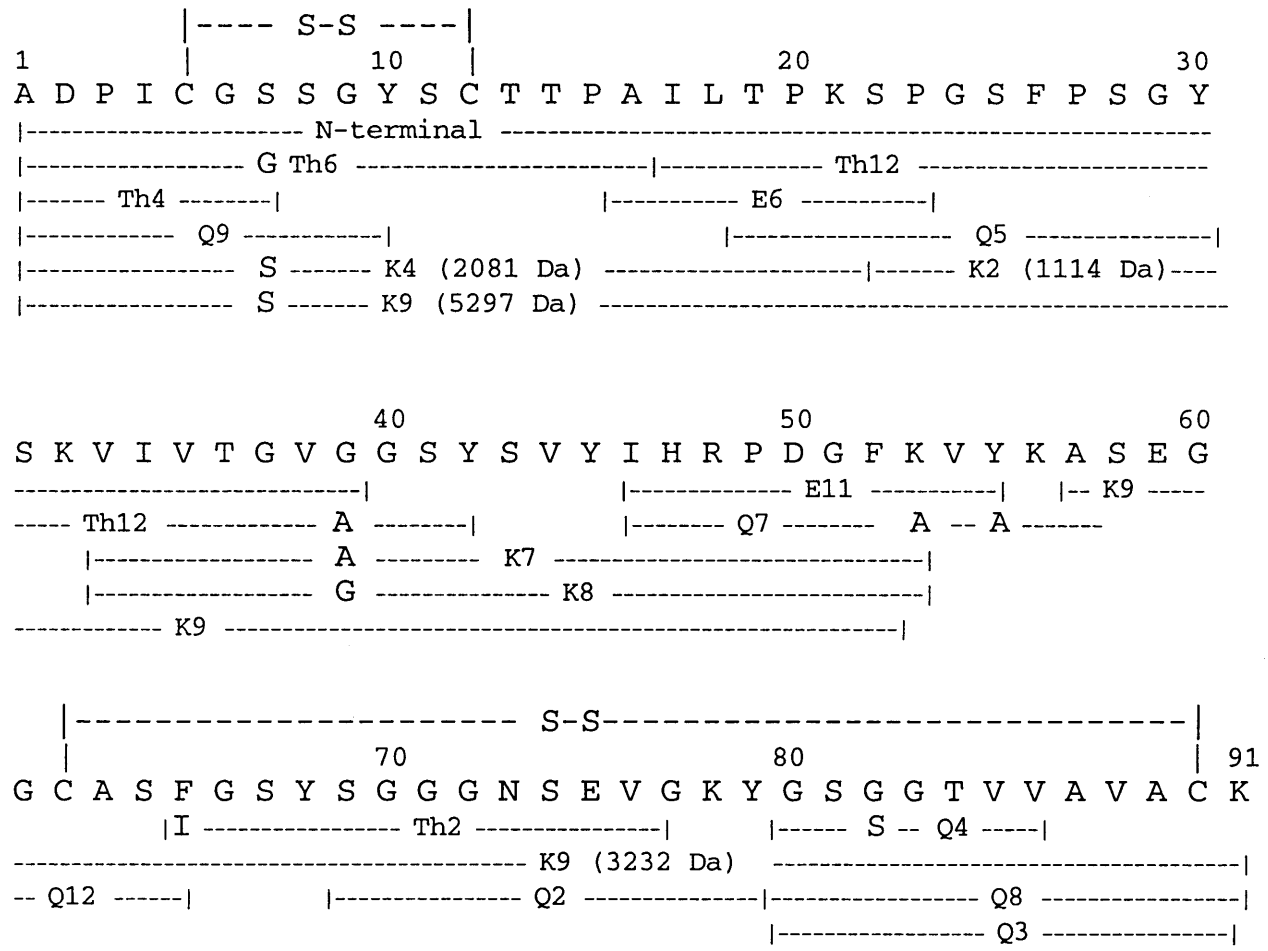

Figure 2. The amino acid sequence of BTL. Peptides obtained by reverse-phase HPLC after degradation of purified BT isolectin mixture with endoproteinase Lys-C, chymotrypsin, elastase, and thermolysin, are denoted K-, Q-, E-, and T-, respectively. Amino acid sequence heterogeneities found at positions 7, 39, 53, 55, 65 and 82 are indicated. S-S indicates the location of a disulphide bond.

from the Dirección General de Investigación Científica y Técnica, Madrid, Spain.

1 Dostálová Z., Calvete J. J., Sanz L. and Töpfer-Petersen E. (1995) Boar spermadhesin AWN-1: oligosaccharide and zone pellucida binding characteristics. Eur. J. Biochem. **230:** 329–336
2 Kasai K.-i. and Hirabayashi J. (1996) Galectins: a family of animal lectins that decipher glycocodes. J. Biochem. **119:** 1–8
3 Gabius H.-J. (1997) Animal lectins. Eur. J. Biochem. **243:** 543–576
4 Damme E. J. M. van, 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
5 Beisel H.-G., Kawabata S.-i., Iwanaga S., Huber R. and Bode W. (1999) Tachylectin-2: crystal structure of a specific GlcNAc/GalNAc-binding lectin involved in the innate immunity host defense of the Japanese horseshoe crab *Tachypleus tridentatus*. EMBO J. **18:** 2313–2322
6 Van Driessche, E., Rougé, P., Beeckmans, S. and Bøg-Hansen, T. C. (eds) (1996) Lectins – Biology, Biochemistry, Clinical Biochemistry, vol. 11, Textop, Denmark
7 Gabius, H.-J. and Gabius S. (eds) (1997) Glycoscience: Status and Perspectives, Chapman & Hall, Weinheim
8 Lys H. and Sharon N. (1998) Lectins: carbohydrate specific proteins that mediate cellular recognition. Chem. Rev. **2:** 637–674
9 Elgavish S. and Shaanan B. (1997) Lectin-carbohydrate interactions: different folds, common recognition principles. Trends Biochem. Sci. **22:** 462–467
10 Loris R., Hamelryck T., Bouckaert J. and Wyns L. (1998) Legume lectin structure. Biochim. Biophys. Acta **1383:** 9–36
11 Bouckaert J., Hamelryck T., Wyns L. and Loris R. (1999) Novel structures of plant lectins and their complexes with carbohydrates. Curr. Opin. Struct. Biol. **9:** 572–577
12 Rini J. M. and Lobsanov Y. D. (1999) New animal lectin structures. Curr. Opin. Struct. Biol. **9:** 578–584
13 Drickamer K. (1999) C-type lectin-like domains. Curr. Opin. Struct. Biol. **9:** 585–590
14 Boyd W. C., Almodovar L. R. and Boyd L. G. (1966) Agglutinins in marine algae for human erythrocytes. Transfusion **6:** 82–83
15 Rogers D. J. and Hori K. (1993) Marine algal lectins: new developments. Hydrobiologia **260/261:** 589–593

16 Hori K., Miyasawa K. and Ito K. (1990) Some common properties of lectins from marine algae. Hydrobiologia **204/205:** 561–566
17 Sampaio A. H., Rogers D. J. and Barwell C. J. (1996) Characterisation of a new lectin from the green marine alga *Ulva laetevirens*. In: Lectins – Biology, Biochemistry, Clinical Biochemistry, vol. 11, pp. 96–100, Van Driessche E., Rougé P., Beeckmans S. and Bøg-Hansen T. C. (eds) Textop, Denmark
18 Sampaio A. H., Rogers D. J. and Barwell C. J. (1998) Isolation and characterisation of the lectin from the green marine alga *Ulva lactuca* L. Bot. Mar. **41:** 427–433
19 Ainonz I. L., Sampaio A. H., Freitas A. L. P., Benevides N. M. B. and Mapurunga S. (1995) Comparative study on hemagglutinins from the red algae *Bryothamnion seaforthii* and *Bryothamnion triquetrum*. Rev. Bras. Fisiol. Veg. **7:** 15–19
20 Ainouz I. L., Sampaio A. H., Benevides M. M. B., Freitas A. L. P., Costa F. H. F., Carvalho M. C. et al. (1992) Agglutination of enzyme treated erythrocytes by Brazilian marine algae. Bot. Mar. **35:** 475–479
21 Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W. et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402
22 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
23 Sampaio A. H., Rogers D. J., Barwell C. J., Saker-Sampaio S., Costa F. H. F. and Ramos M. V. (1998) A new procedure and further characterisation of the lectin from the red marine alga *Ptilota serrata*. J. Appl. Phycol. **10:** 539–546