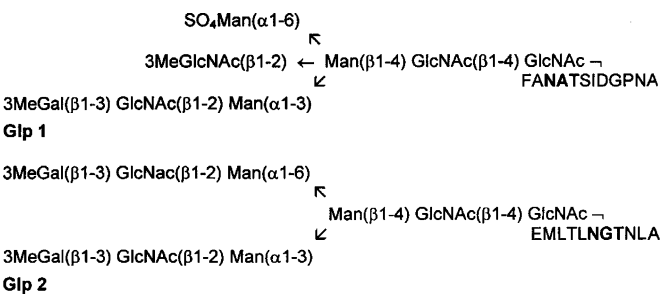
Characterization of the carbohydrate moieties of the functional unit RvH_1-a of *Rapana venosa* haemocyanin using HPLC/electrospray ionization MS and glycosidase digestion

Pavlina DOLASHKA-ANGELOVA*[1], Alexander BECK†, Alexandar DOLASHKI‡, Mariano BELTRAMINI§, Stefan STEVANOVIC‖, Benedetto SALVATO§ and Wolfgang VOELTER‡

*Institute of Organic Chemistry, Bulgarian Academy of Sciences, G. Bonchev 9, Sofia 1113, Bulgaria, †Klinisch-chemisches Zentrallaboratorium der Universitätskliniken, Abteilung Innere Medizin IV, Universität Tübingen, Otfried-Müller-Straße 10, 72076 Tübingen, Germany, ‡Abteilung für Physikalische Biochemie des Physiologisch-chemischen Instituts der Universität Tübingen, Hoppe-Seyler-Straße 4, 72076 Tübingen, Germany, §Department of Biology and CNR Institute for Biomedical Technologies, Section of Padova, University of Padova, Via Ugo Bassi 58/B, I-35131 Padova, Italy, and ‖Department of Immunology, Institute for Cell Biology, University of Tübingen, Auf der Morgenstelle 15, D-72076 Tübingen, Germany

The primary structures of two biantennary *N*-glycans of the glycoprotein *Rapana venosa* (marine snail) haemocyanin were determined. Two different structural subunits have been found in *R. venosa* haemocyanin: RvH_1 and RvH_2. The carbohydrate content of the N-terminal functional unit RvH_1-a of RvH_1 was studied and compared with the N-terminal functional unit RvH_2-a of RvH_2. Oligosaccharide fragments were released from the glycoprotein by Smith degradation of a haemocyanin pronase digest and separated on a Superdex 300 column. The glycopeptide fragments, giving a positive reaction for the orcinol/H_2SO_4 method, were separated by HPLC. In order to determine the linked sugar chains to the hinge glycopeptides isolated from the functional unit RvH_1-a, several techniques were applied, including capillary electrophoresis, matrix-assisted laser desorption ionization-MS and electrospray ionization-MS in combination with glycosidase digestion. On the basis of these results and amino acid sequence analysis, we concluded that the functional unit RvH_1-a contains 7 % oligosaccharides N-glycosidically attached to Asn^{262} and Asn^{401}, and the following structures were suggested:



Key words: electrospray ionization MS, glycosylation sites, haemocyanin, matrix-assisted laser-desorption ionization (MALDI)-MS, N-linked oligosaccharides.

INTRODUCTION

Haemocyanins (Hcs) are high-molecular-mass (4.5×10^2 to 9×10^3 kDa) copper-containing oxygen-transporting proteins, freely dissolved in the haemolymph of several arthropodan and molluscan species [1–5]. There are two different kinds of Hcs, one found in arthropods and the other isolated from molluscs. Both of them use a copper pair for oxygen binding, which is part of the active site, but their subunit sizes are different as well as their subunit organization. There is low overall sequence similarity between the two classes of Hcs, although the similarity is significantly higher in the active-site region [6,7], and peculiar features of the molecular organization point to different evolutionary pathways [8].

Most Hcs are glycoproteins, although there are large differences in their carbohydrate contents and their monosaccharide composition, and both O-linked and N-linked oligosaccharides were identified. The monosaccharide compositions and the carbohydrate contents of Hcs from various arthropodan species usually range between 0.1 and 2 %, and D-mannose (Man) and *N*-acetyl-D-galactosamine (GalNAc) are the most abundant residues [9,10]. For the Hc from the centipede *Scutigera coleoptrata*, an exceptionally high carbohydrate content (4.9 %) was found [11]. Recently, we have reported [12] that in the hexameric Hc from the crab *Carcinus aestuarii*, where the carbohydrate moiety accounts for 1.6 % of total protein mass, only one subunit is specifically involved in carbohydrate binding with a carbohydrate content of 6.3 % (w/w), which is higher in comparison with the carbohydrate content identified for other arthropodan structural subunits. Three consensus sequences for O-glycosylation and one for N-glycosylation were found by sequencing the glycopeptides isolated after tryptic digestion of the subunit [12]. Since this subunit is not able to re-associate into hexamers after dissociation of the native protein, it is tempting to correlate such characteristics with the presence of the carbohydrate chains. Therefore the appropriate packing of the carbohydrate fraction could represent an important factor to be considered for interpreting the irreversibility of re-association phenomena.

Carbohydrate contents of molluscan Hcs have been studied from proteins isolated from the terrestrial snail *Helix pomatia* [13], the freshwater snail *Lymnaea stagnalis* [14], the marine gastropod *Rapana thomasiana* [15] and the keyhole limpet *Megathura crenulata* [16]. A relatively high carbohydrate content, between 2 and 9 % (w/w), is typical for these molluscan Hcs, and as monosaccharides xylose, fucose (Fuc), 3-*O*-methyl-D-galactose (3MeGal), Man, D-galactose (Gal), GalNAc and *N*-acetyl-D-glucosamine (GlcNAc) were determined. The carbohydrate moiety of molluscan Hcs has recently received particular interest

Abbreviations used: ESI-MS, electrospray ionization MS; Fuc, fucose; GalNAc, *N*-acetyl-D-galactosamine; GlcNAc, *N*-acetyl-D-glucosamine; Glp1, glycoprotein 1; Glp2, glycoprotein 2; Hc, haemocyanin; MALDI-MS, matrix-assisted laser-desorption ionization MS; Man, D-mannose; 3MeGal, 3-*O*-methyl-D-galactose; PNGase F, peptide N-glycosidase F; RvH1 and RvH2, structural subunits of *Rapana venosa* Hc.

[1] To whom correspondence should be addressed (e-mail pda54@yahoo.com).

for their immunostimulatory properties [17]. In the present study, we have focused on Hc isolated from the marine snail *R. venosa* (RvH) referred to previously as *R. thomasiana* grosse [18,19]. In a previous study on the glycosylation of *R. thomasiana* Hc [15], only the ratios of the oligosaccharides were determined; however, no information about the sequence of the monomers and their linkage sites was reported [15]. In the present study, we concentrated on the carbohydrate structures of the N-terminal functional unit RvH_1-a of the structural subunit RvH_1 of *R. venosa* Hc.

EXPERIMENTAL

Isolation of the functional unit RvH1-a from RvH_1 of *R. venosa* Hc

Native Hc was purified from the haemolymph of *R. venosa* Hc as described previously [18–20]. Subunit RvH_1, referred to previously as *Rapana* haemocyanin structural subunit 1 ('RHSS1'), was eluted as the first peak from an ion-exchange chromatography column with a 0–0.5 M NaCl gradient in 50 mM Tris/HCl buffer (pH 8.2). Subsequently, 200 mg of the subunit was treated with trypsin (trypsin/Hc ratio, 1:400) for 1 h at room temperature in 20 mM NH_4HCO_3 buffer (pH 8.2). The tryptic hydrolysate was separated on a Sephadex G-150 column, eluted with the same buffer at a flow rate of 1 ml · min^{-1}. The last eluting peak fraction, containing RvH_1-a, was loaded on to a Mono Q 10/10 column (FPLC) equilibrated with 50 mM Tris/HCl buffer (pH 8.2), and the functional unit was eluted with a linear gradient (0–0.5 M NaCl in 60 min) at a flow rate of 1 ml · min^{-1} and desalted on a Sephadex G-25 column with water.

Preparation of copper-free Hc

To prepare the starting material suitable for glycopeptide analysis, the apo-protein was prepared. Freeze-dried portions of RvH1-a were dialysed overnight against 50 mM Tris/HCl buffer (pH 8.2) containing 10 mM KCN, with three changes of buffer.

Carbohydrate determination and protein digestion

RvH1-a (8 mg) was dissolved in 1 ml of 0.4 M Tris/HCl buffer (pH 8.6) containing 6 M guanidine/HCl and 0.2 M EDTA. To cleave the disulphide bonds, 20 μl of 98 % (v/v) 2-mercaptoethanol was added with stirring. After heating at 50 °C for 4 h, 3 μl of 4-vinylpyridine was added and the reaction mixture allowed to stand for 3 h at room temperature. The reaction was terminated by the addition of 50 μl of 2.0 M acetic acid. The sample was dissolved in 200 μl of 0.1 M ammonium bicarbonate buffer (pH 9.0), and 50 μl of the trypsin solution (bovine pancreas; Hc/trypsin ratio, 50:1) was added and the reaction mixture incubated at room temperature for 3 h. The glycopeptide mixture was separated on a Superdex 300 gel-filtration column (2 cm × 30 cm), and the fractions were eluted with water at a flow rate of 1 ml · min^{-1}. Each chromatographic peak fraction was checked for carbohydrates using the orcinol/H_2SO_4 test [12,21]. The only peak fraction giving a positive reaction was fractionated further by reverse-phase HPLC using a Nucleosil 7 C18 column (250 mm × 10 mm; Macherey-Nagel, Düren, Germany). For elution, a linear gradient of 5 % solvent A (0.1 % trifluoroacetic acid in water) and 100 % solvent B (0.085 % trifluoroacetic acid in acetonitrile) within 70 min at a flow rate of 1 ml · min^{-1} was used. The HPLC fractions, detected at a wavelength of 206 nm, were collected, freeze-dried and analysed for carbohydrates with orcinol/H_2SO_4 on silica-gel plates.

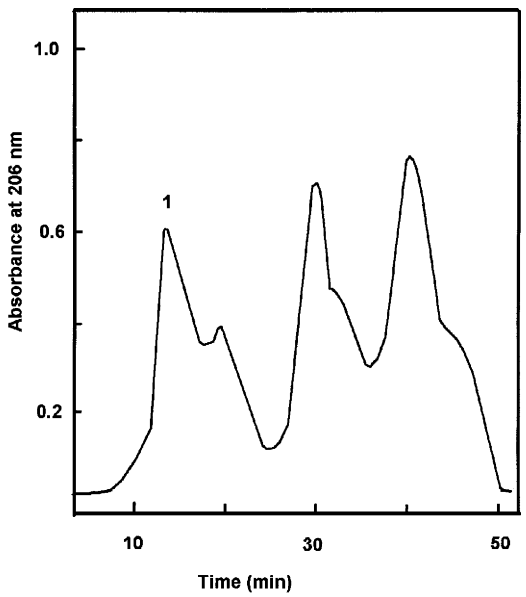


Figure 1 Separation of the tryptic digest obtained from the functional unit RvH_1-a

Gel filtration of the tryptic digest was separated on a Superdex 300 column (2 cm × 30 cm) at a flow rate of 1 ml · min^{-1}. Peak fraction 1 gave a positive reaction in the orcinol/H_2SO_4 test and was processed further.

Glycoprotein/peptide staining on silica-gel plates

The freeze-dried peptides were dissolved in water and 2–4 μl was transferred to the plate, taking care to restrict the size of the spot to 2–3 mm in diameter, and air dried. The plate was sprayed with orcinol/H_2SO_4 and heated for 20 min at 100 °C [21].

Amino acid sequence analysis

Amino acid sequence analysis was performed for the peptides that were positive in the orcinol test. The fractions were dried and, after dissolving in 40 % methanol/1 % formic acid (v/v), subjected to automated Edman N-terminal sequencing (Procise 494A Pulsed Liquid Protein Sequencer; Applied Biosystems GmbH, Weiterstadt, Germany).

Enzymic digestion of glycopeptides

The fractions giving a positive test for carbohydrates, covering an amino acid sequence fragment Asn-Xaa-Ser/Thr, were N-deglycosylated by peptide N-glycosidase F (PNGase F) from *Flavobacterium meningosepticum* (Calbiochem). The glycopeptides were dissolved in 50 mM Tris/HCl buffer (pH 7.0) and 5 μl of PNGase F (1.5 units · ml^{-1}) was added [22,23]. After incubation for 24 h at room temperature, the samples were analysed by matrix-assisted laser-desorption ionization (MALDI)-MS as described below.

Enzymic digestion of glycopeptides was performed in a total volume of 50 μl in 0.7 ml Eppendorf tubes in 50 mM Tris/HCl buffer (pH 7.0). The glycosidases PNGase F, β1-2,3,4,6-*N*-acetylglucosaminidase, α1-2,3-mannosidase, α1-2,3,6-mannosidase (all from recombinant *Escherichia coli*) and β1-3,4,6-galactosidase (from bovine testes) were used (Calbiochem) in sequence, and the reactions were started with the addition of 5 μl of these enzymes. After incubation for 24 h at 37 °C, the samples were

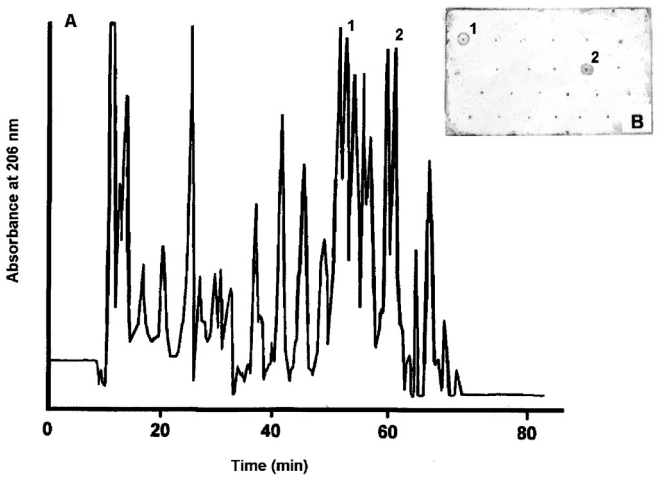


Figure 2 HPLC profile of peak 1

(A) Peak fraction 1 obtained from Figure 1 was applied on to a Nucleosil RP18 (C18, 100 mm length × 2.1 mm diameter) column and eluted with a linear gradient of 5 % solvent A (0.1 % trifluoroacetic acid in water) and 100 % solvent B (0.085 % trifluoroacetic acid in acetonitrile) in 70 min at a flow rate of 1 ml · min^{-1}. Proteins were detected at a wavelength of 206 nm. (**B**) Orcinol/H_2SO_4 test of (glyco) peptides eluted by HPLC and applied on to a silica-gel plate.

analysed by capillary electrophoresis and electrospray ionization MS (ESI-MS).

MALDI-MS analysis of the glycopeptides

The glycosylated and deglycosylated peptides were analysed by MALDI-MS by means of a Kratos MALDI III equipment (Shimadzu). The glycopeptides were dissolved in 0.1 % trifluoracetic acid (v/v) and applied on to the target. The matrix α-cyano-4-hydroxycinnamic acid in 70 % acetonitrile/H_2O (70:30, v/v) was used.

ESI-MS

ESI-MS was performed with a triple quadrupole mass spectrometer (TSQ700; Finnigan MAT, Bremen, Germany) equipped with an ESI-Finnigan electrospray ion source. The targeted fractions were dried at room temperature using a speed vacuum evaporator (Savant, New York, NY, U.S.A.) and dissolved in 50 % water/methanol (v/v). After centrifugation (2 min, 2000 ***g***), the supernatant was infused into the ESI-MS via a Harvard syringe pump (5 μl/min). Mass spectra were acquired in positive-ion mode. Q1 was scanned over a mass range of *m*/*z* 400–2000 for 3 s or *m*/*z* 400–3000 for 3.5 s.

Capillary electrophoresis

All separations were performed on a BioFocus 3000 CE instrument (Bio-Rad, Munich, Germany). Separations were performed on a 50 cm × 50 μm (internal diameter) fused silica capillary (Grom, Herrenberg, Germany). The samples were dissolved in 50 μl of running buffer [50 mM sodium phosphate buffer (pH 2.5)] diluted with water (10:1, v/v) and introduced by pressure injection. All electrophoretic separations were carried out at 25 kV constant voltage and the capillary temperature was maintained at 25 °C. The peptides were detected by absorption at $\lambda = 214$ nm.

RESULTS AND DISCUSSION

Isolation of glycopeptides from the functional unit RvH_1-a

Rapana Hc is a glycoprotein, and a carbohydrate content of 8.9 % was determined for the native molecule and 12.8 % and 4.4 % for RvH_1 and RvH_2 respectively [15]. The oligosaccharide content is more abundant in the N-terminal functional unit RvH1-a (7 %) of RvH_1 than in N-terminal functional unit RvH2-a (5.1 %) of RvH_2. The sugar content of the marine snail *R. venosa* Hc is similar to that of the protein from *H. pomatia* (8.25 %) [13] and approx. 3 times higher compared with *L. stagnalis* Hc (3.01 %) [14].

To confirm that the RvH_1-a is indeed glycosylated, a screen on its carbohydrate content was performed using the orcinol method. The strategy used to analyse the carbohydrate portion of RvH_1-a was to prepare a tryptic hydrolysate of the functional unit and separate its fragments, first by gel filtration and then by HPLC. Fractions giving a positive orcinol colour reaction were further subjected to amino acid and carbohydrate sequence determination. Figure 1 shows the separation profile of the tryptic digestion products of functional unit RvH_1-a using a

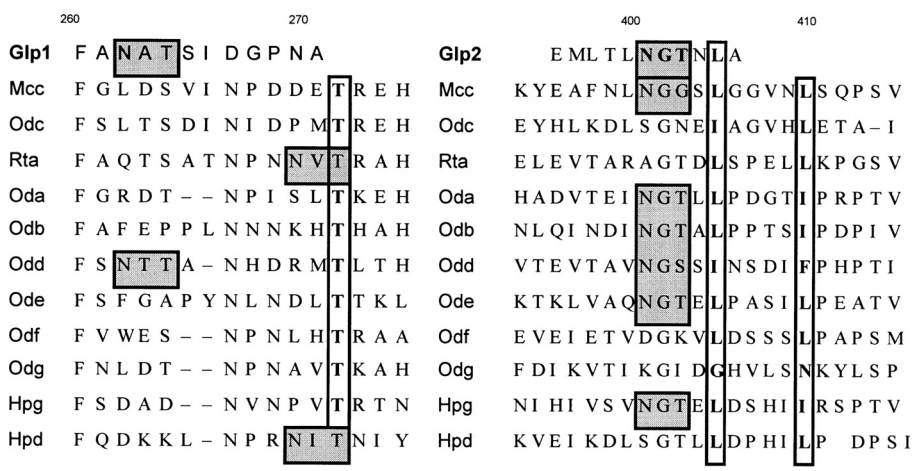


Figure 3 Sequence alignments of the regions with suggested N-linked sites of functional units from different molluscan Hcs

Glp1 and Glp2 were obtained from Figure 2 and their sequences obtained after removal of the carbohydrate chains with PNGase F. Glp1 and Glp2 are aligned with the following molluscan Hcs: keyhole limpet *M. crenulata* (Mcc) [16], *O. dofleini* (Oda, Odb, Odc, Odd, Ode, Odf and Odg) [27], *R. thomasiana* (Rta) [28] and *H. pomatia* (Hpd and Hpg) [29].

Table 1 Monosaccharide composition of Fus RvH1-a from *Rapana venosa* Hc

Fraction	Amino acid sequence of the peptide	Molecular mass (Da)		
		Glycopeptide	Peptide	Oligosaccharide
1	FA**NAT**SIDGPNA	2786 $[M+Na]^+ = 2763$	1177.0* 1177.5†	1609 $[M+Na]^+ = 1586$
2	EMLTL**NGT**NLA	2846.2 $[M+H]^+ = 2828$	1175.7* 1175.6†	1653 $[M+H]^+ = 1652$

* Molecular mass determined by MALDI-MS.
† Molecular mass calculated by amino acid sequence.

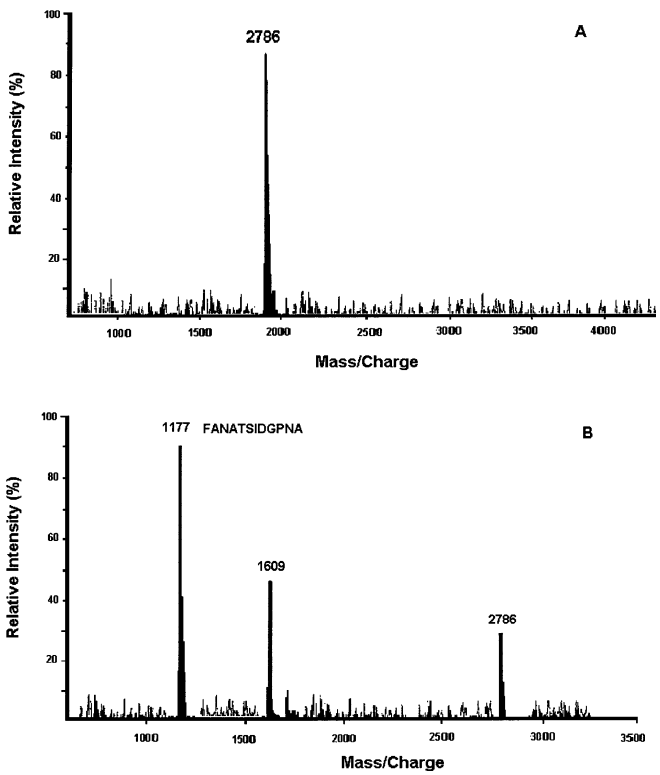


Figure 4 MALDI-MS of Glp1 before (A) and after (B) glycosidase digestion

Glp1 from functional unit RvH_1-a was isolated by HPLC, as shown in Figure 2, and investigated by MALDI-MS before (**A**) and after (**B**) treatment with PNGase F.

Superdex 300 column. The material eluted in peak fraction 1 (Figure 1) was pooled and peptides were separated by reverse-phase HPLC. Each peak was collected, vacuum-concentrated and tested for carbohydrates on a silica-gel plate with orcinol/H_2SO_4. Two fractions, 1 and 2 (Figure 2), gave a positive reaction for carbohydrates (Figure 2B) and were studied further using ESI-MS [24,25] and MALDI-MS [26].

Glycopeptide sequences

The two glycopeptides (Glp1 and Glp2) were automatically sequenced after removal of the carbohydrate chains with PNGase F and their sequences are shown in Figure 3. These sequences are compared with specific sequences of other molluscan Hcs containing carbohydrate linkages. Glp1 consists of 12 amino acid residues with a molecular mass calculated from the sequence of 1177.5 Da. LALIGN and Fasta programmes were used to analyse the alignment of Glp1 and Glp2 with the amino acid sequences of other molluscan Hcs. The amino acid sequence of Glp1 [Phe-Ala-**Asn**-**Ala**-**Thr**-Ser-Ile-Asp-Gly-Pro-Asn-Ala (FA**NAT**SIDGPNA), where residues in bold represent a consensus sequence for an N-linked glycosylation site] overlaps well with positions 260–271 of the functional units of molluscan Hcs from *M. crenulata* (Mcc) [16], *O. dofleini* (Oda–Odg) [27], *R. thomasiana* (Rta) [28] and *H. pomacia* (Hpd and Hpg) [29]. The presence of a fully conserved phenyalanine residue at position 260 and an asparagine/aspartic acid residue at position 267 allows the alignment with the other reported sequences, despite the rather high variability in this region. From all sequences listed in Figure 3, a glycosylation site was only observed in the Odd Hc fragment. The same considerations apply for Glp2 [Glu-Met-Leu-Thr-Leu-**Asn**-**Gly**-**Thr**-Asn-Leu-Ala (EMLTL**NGT**NLA), where residues in bold represent a consensus sequence for an N-linked glycosylation site] in Figure 3, where Gly^{402}, Leu^{405} and Leu^{410} are conserved residues. A sequence containing 11 amino acid residues with a calculated molecular mass of 1175.6 Da was found for Glp2. This glycopeptide contains the Asn-Gly-Thr sequence that is highly conserved among molluscan Hcs, including *M. crenulata* (Mcc) [16], *O. dofleini* (Oda–Ode) [27] and *H. pomatia* (Hpd) [29], in the 401–403 sequence region (Asn-Gly-Thr/Ser), and Glu^{396} is found in six out of 12 residues and Leu^{405} is found in nine out of 12 cases. Thus two putative glycosylation sites are found in the sequence of RvH1-a, with consensus sequences for N-linked carbohydrate oligosaccharides.

Composition of the carbohydrate portion of Glp1

Glp1 was analysed by MALDI-MS before and after treatment with PNGase F (Table 1 and Figure 4). In the MALDI mass spectrum acquired before PNGase F treatment, only one molecular ion at *m/z* 2786 was detected, caused by the intact Glp1 sodium adduct ion $[Glp1 + Na]^+$ (Figure 4A). In contrast, after PNGase-F treatment, three signals were observed in the MALDI mass spectrum (Figure 4B). The signal at *m/z* 2786 is related to the intact Glp1 (see Figure 4A), whereas the signal at *m/z* 1609 $[M + Na]^+$ is attributed to an oligosaccharide. The first signal (*m/z* 1177) corresponded to the deglycosylated peptide, as demonstrated by the mass value calculated from the amino acid sequence (1177.5 Da; Table 1).

To determine the oligosaccharide sequence, different glycolytic enzymes (β1-2,3,4,6-GlcNAcase, α1-2,3-mannosidase and α1-2,3,6-mannosidase and β1-3,4,6-galactosidase) were added to Glp1. The mass (1586 Da; Table 1) of this carbohydrate moiety would account for an oligosaccharide chain containing $(SO_4)MeGalGlcNAc_4Man_3$ connected to the peptide. Depending on the specificity of the individual enzymes, the different linkages can be identified by recording the pattern of molecular masses resulting from the digestion with the pool of glycosidases. After incubation for 24 h at 37 °C, the sample was first analysed by capillary electrophoresis (Figure 5, inset). Four different peaks were detected, indicating that the treatment causes heterogeneity resulting from different cleavage sites.

In parallel, Glp 1 treated with the different glycolytic enzymes was analysed by ESI-MS (Figure 5). The mass spectrum was interpreted on the basis of the data obtained from MALDI-MS (Figure 4), amino acid sequence (Figure 3) and the known specificity of the glycosidases.

The signal observed at *m/z* 1747.0 may be attributed to the peptide Phe-Ala-Asn-Ala-Thr-Ser-Ile-Asp-Gly-Pro-Asn-Ala

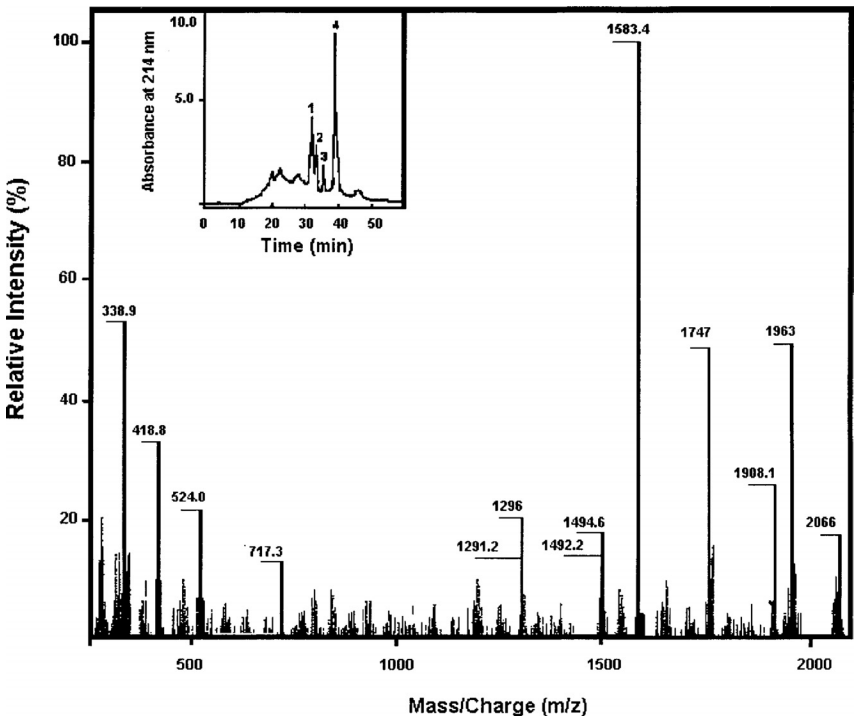


Figure 5 ESI-MS and capillary electrophoresis of Glp1 following glycosidase digestion

Positive-ion ESI mass spectrum of Glp1 after enzymic cleavage with β1-2,3,4,6-GlcNAcase, α1-2,3,6-mannosidase and α1-2,3-mannosidase. Inset, separation of Glp1, isolated by HPLC from Figure 2, by capillary electrophoresis after treatment with different glycosidases. Conditions of capillary electrophoresis was carried out using an uncoated fused-silica capillary column (50 cm $\times$ 50 μm internal diameter) at 25 kV in 50 mM sodium phosphate buffer (pH 2.5).

Table 2 Proposed structures of the carbohydrate chains of Glp1 and Glp2

The carbohydrate chains were calculated on basis of observed $[M + H]^+$ signals in the ESI mass spectra from Figures 5 and 7. The enzymes used for carbohydrate cleavage are shown. **P** represents the peptide with sequence FA**NAT**SIDGPNA (1177 Da).

No	Enzymes	$[M+H]^+$ (m/z)	Structure
	Glycopeptide 1		
A	α1-2,3- and α1-2,3,6-mannosidase and β1-2,3,4,6-N-GlcNAcase	1747	← Man(β1-4) GlcNAc(β1-4) GlcNAc ¬P
B	α1-2,3-mannosidase and β1-2,3,4,6-N-GlcNAcase	1908	Man(α1-6) ↖ ← Man(β1-4) GlcNAc(β1-4) GlcNAc ¬P
C	α1-2,3,6-mannosidase and α1-2,3-mannosidase	1953	3MeGlcNAc(β1-2) ←Man(β1-4) GlcNAc(β1-4) GlcNAc¬P
D	β1-2,3,4,6-N-GlcNAcase	1583	-GlcNAc(β1-4) GlcNAc¬P
	Glycopeptide 2		
E	α1-2,3-mannosidase, β1-2,3,4,6-N-GlcNAcase and PNGase F	731	Man(α1-6) ↖ ← Man(β1-4) GlcNAc(β1-4) GlcNAc

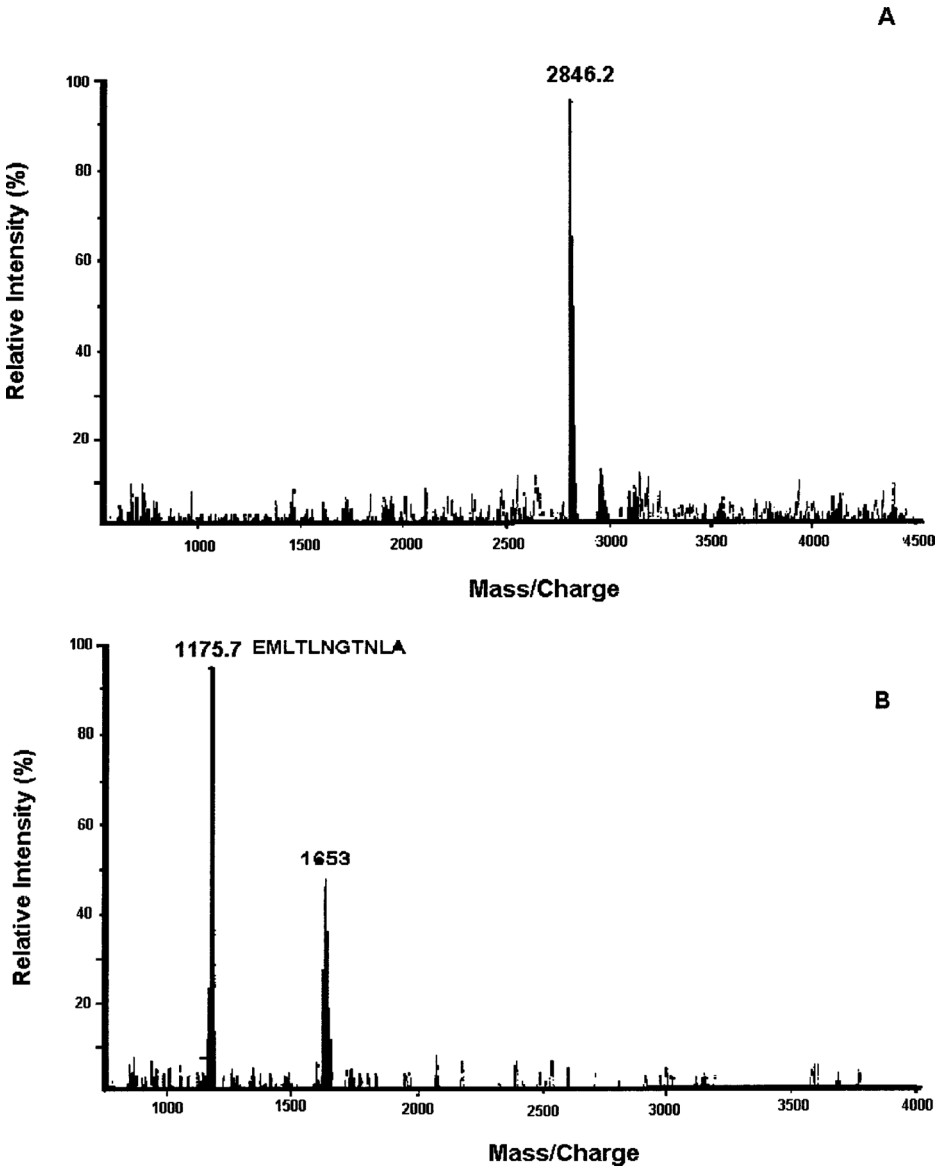


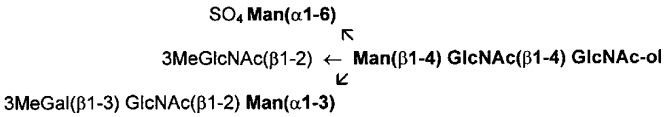
Figure 6 MALDI-MS of Glp2 before (A) and after (B) glycosidase digestion

A) Glp2 from the functional unit RvH_1-a was isolated by HPLC, as shown in Figure 2, and investigated by MALDI-MS before (**A**) and after (**B**) treatment with PNGase F.

(1177.5 Da) containing two GlcNAc and one Man(β1-4) residues (568 Da), as expected from the specificity of α1-2,3,6-, α1-2,3-mannosidases and β1-2,3,4,6-GlcNAcase in removing the carbohydrates attached to Man(α1-6) and Man(α1-3), connected to Man(β1-4) (Table 2A). The signal at *m/z* 1908.1 may result from the glycopeptides containing two GlcNAc, one Man(β1-4) and one Man(α1-6) residue (730 Da), as expected from the specificity of α1-2,3-mannosidase, cleaving only Man(α1-3), and β1-2,3,4,6-GlcNAcase removing 3MeGal and GlcNAc(β1-2) from Man(α1-6) (Table 2B). This loss of 856 Da compared with the intact glycopeptide suggests the following residues were removed: MeGal, GlcNAc, Man, MeGlcNAc and SO_4.

The peak at *m/z* 1953 corresponds to the carbohydrate fragment containing two GlcNAc, one Man(β1-4) and possibly 3MeGlcNAc(β1-2), resulting from the activity of α1-2,3,6- and α1-2,3-mannosidases removing carbohydrates attached to Man(α1-6) and Man(α1-3) connected to Man(β1-4) (Table 2C). The signal observed at *m/z* 1583.4 can be attributed to the peptide Phe-Ala-Asn-Ala-Thr-Ser-Ile-Asp-Gly-Pro-Asn-Ala (FANATSIDGPNA; 1177.5 Da) containing two GlcNAc groups (Table 2D).

Based on these data, the carbohydrate structure for Glp1 from functional unit RvH1-a is proposed as shown below:



Carbohydrate content of Glp2

The use of specific glycosidases, as described above, and the combination of changes in glycopeptide mass after each digestion step also allow a sequence for Glp2 to be suggested. Only one

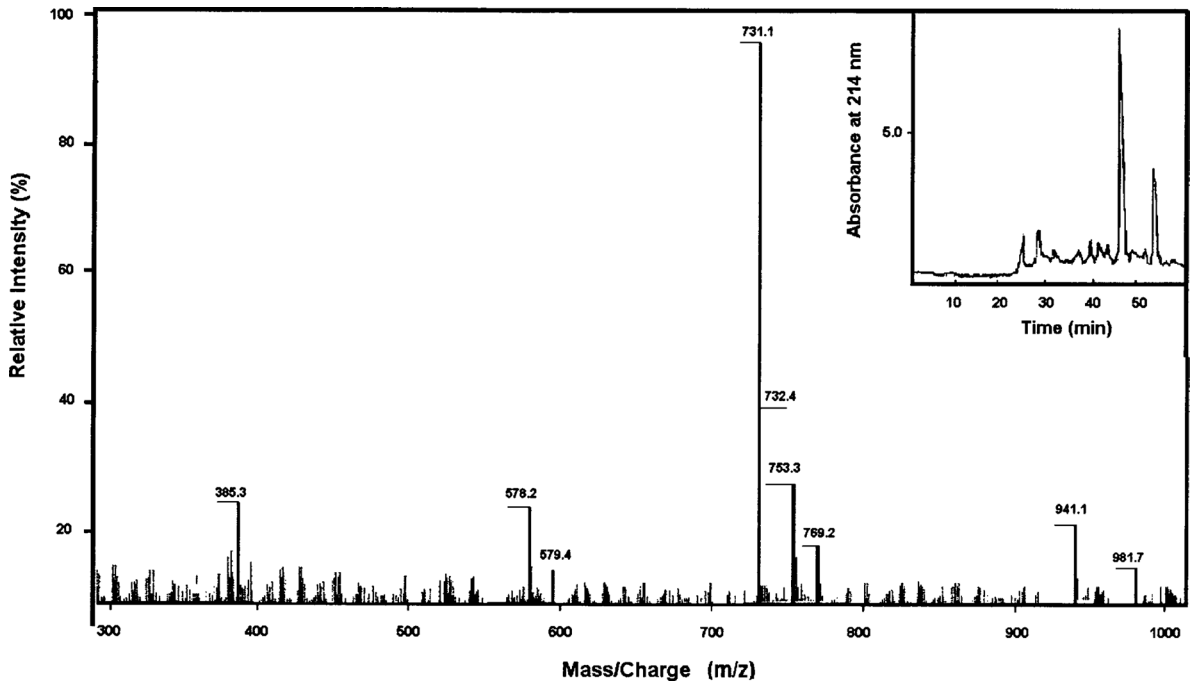
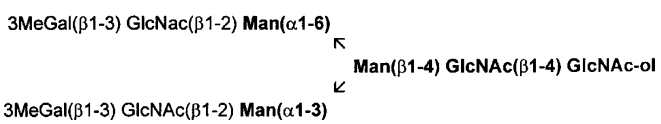


Figure 7 ESI-MS and capillary electrophoresis of Glp2 following glycosidase digestion

Positive-ion ESI mass spectrum of Glp2 after enzymic cleavage with β1-2,3,4,6-GlcNAcase, α1-2,3,6-mannosidase and α1-2,3-mannosidase. Inset, separation of Glp2, isolated by HPLC from Figure 2, by capillary electrophoresis after treatment with different glycosidases. Conditions of capillary electrophoresis was carried out using an uncoated fused-silica capillary column (50 cm $\times$ 50 μm internal diameter) at 25 kV in 50 mM sodium phosphate buffer (pH 2.5).

peak of 2846 Da $[M + 2H]^{2+}$ was observed in the MALDI mass spectrum (Figure 6A). As shown in Figure 6(B), two peaks were present in the MALDI-MS spectrum of the oligopeptide after cleavage with PNGase F. The signal observed at *m/z* 1175 can be attributed to the peptide Glu-Met-Leu-Thr-Leu-Asn-Gly-Thr-Asn-Leu-Ala (EMLTLNGTNLA) with the molecular mass of 1175.6 Da, calculated on the basis of its amino acid sequence (Table 1). Thus the peak at *m/z* 1653 can be tentatively assigned to an oligosaccharide with structure $3MeGal_2GlcNAc_4Man_3$. The presence of a consensus sequence for one N-linked glycosylation site (Asn-Gly-Thr) of the peptide chain indicates that the oligosaccharide is connected to Asp[401] via GlcNAc. Two peaks were also separated by capillary electrophoresis after treatment of Glp2 with PNGase F only, corresponding to the peptide and the carbohydrate chain (Figure 7 inset). Glp2 was treated with β-3,4,6-galactosidase, β1-2,3,4,6-GlcNAcase and α1-2,3-mannosidase and subsequently analysed by ESI-MS (Figure 7). Because Glp2 was cleaved with PNGase F before glycosidase treatment, the N-linked carbohydrate was removed from the glycopeptide. Therefore the signal at *m/z* 731 in the ESI mass spectrum is tentatively assigned to $Man_2GlcNAc_2$ (Table 2E).

Based on the cleavage with specific glycosidases, followed by observed MS data, we suggest that the carbohydrate structure of Glp2 from functional unit RvH_1-a could be assigned as follows:



Our results are consistent with the presence of a common trimannosyl-*N*,*N*′-diacetylchitobiose core Man(α1-6) [Man(α1-3)]Man(β1-4)GlcNAc(β1-4)GlcNAc-ol and different antennae are attached to the α-Man residues [13,14]. This structural principle also applies to Hcs of different molluscs as well as to the functional units RvH_1-a (the present study) and RvH_2-e [28]. The recent characterization carried out on the carbohydrate chains of all functional units of *H. pomatia* Hc [13] identified primary structures of 21 novel monoantennary and diantennary N-glycans of the glycoprotein besides its core element. The oligosaccharide fragments (antennae) were released from the glycoprotein by Smith degradation of an Hc pronase digest and the major antennae were characterized using ^{1}H NMR spectroscopy and fast atom bombardment MS. In the present study, however, the number of carbohydrate chains for each functional unit was not defined, because the whole Hc was used, and also the linkage sites were not identified. In the present study, we show that two N-linkage sites are present within one functional unit and that the two different chains differ in their branching characteristics. From the sequence alignment, we suggest that the presence of two N-linkage sites is characteristic for the functional units of other Hcs as well. As far as the N-terminal functional units of *R. venosa* Hc RvH_1-a and RvH_2-a are concerned, the sequence of the former shows the presence of two consensus sequences: one is found in the same region as Glp 1 (Figure 3), whereas a second one is located upstream close to the N-terminus (Asn-Asp-Ser; residues 32–34 [28]; results not shown in Figure 3). As is obvious from Table 1, the glycosidic linkage site at position 401–403 overlaps well with several other Hcs. These results are in agreement with the X-ray study of RvH2-e [30], identifying two oligosaccharide side chains connected to Asn[127] and Asn[17], although only one

glycopeptide has been so far isolated, but not identified from the same source.

P. D.-A. would like to thank Deutsche Forschungsgemeinschaft (DFG) and Deutscher Akademischer Austausch Dienst (DAAD) for granting a scholarship. This work was supported by grants from North Atlantic Treaty Organization (NATO) (LST.CLG.978560) and from the NCSI (X-1202) of the Ministry of Education and Science, Bulgaria.

REFERENCES

1 Salvato, B. and Beltramini, M. (1990) Hemocyanin: molecular architecture, structure and reactivity of the binuclear copper active site. Life Chem. Rep. **8**, 1–47
2 Markl, J. (1986) Evolution and function of structurally diverse subunits in the respiratory protein hemocyanin from arthropods. Biol. Bull. **171**, 90–115
3 Stoeva, S., Dolashka, P., Pervanova, K., Genov, N. and Voelter, W. (1998) Multidomain structure of the *Rapana thomasiana* (Gastropod) hemocyanin structural subunit RHSS1. Comp. Biochem. Physiol., Part B: Biochem. Mol. Biol. **118**, 927–934
4 Van Holde, K. E. and Miller, K. I. (1995) Hemocyanins. Adv. Protein Chem. **47**, 1–81
5 Herskovits, T. T. (1988) Recent aspects of the subunit organization and dissociation of hemocyanins. Comp. Biochem. Physiol. Part B: Biochem Mol. Biol. **91**, 597–611
6 Lang, W. H. and van Hold, K. E. (1991) Cloning and sequencing of *Octopus dofleini* hemocyanin cDNA: derived sequences of functional units Ode and Odf. Proc. Natl. Acad. Sci. U.S.A. **88**, 244–248
7 Neuteboom, B., Jekel, P. A. and Beintema, J. J. (1992) Primary structure of hemocyanin subunit c from *Panulirus interruptus*. Eur. J. Biochem. **206**, 243–249
8 Van Holde, K. E., Miller, K. I. and Decker, H. (2001) Hemocyanins and invertebrate evolution. J. Biol. Chem. **276**, 19, 15563–15566
9 Van Kuik, J. A., Kamerling, J. P. and Vliegenthart, J. F. G. (1990) Carbohydrate analysis of hemocyanins. In Invertebrate Oxygen Carriers (Preaux, G. and Lontie, R., eds.), pp. 157–163, Leuven University Press, Leuven
10 Markl, J., Schmid, R., Czichos-Tiedt, S. and Linzen, B. (1976) Haemocyanins in spiders, III. Chemical and physical properties of the proteins in *Dugesiella* and *Cupiennius* blood. Hoppe–Seyler's Z. Physiol. Chem. **357**, 1713–1725
11 Waxman, L. (1975) The structure of arthropod and mollusc hemocyanins. J. Biol. Chem. **250**, 3796–3806
12 Dolashka-Angelova, P., Beltramini, M., Dolashki, A., Salvato, B., Hristova, R. and Voelter, W. (2001) Carbohydrate composition of *Carcinus aestuarii* hemocyanin. Arch. Bioch. Biophys. **389**, 153–158
13 Lommerse, J. P. M., Thomas-Oates, J. E., Gielens, C., Preaux, G., Kamerling, J. P. and Vliegenthart, J. F. G. (1997) Primary structure of 21 novel monoantennary and diantennary N-linked carbohydrate chains from αD-hemocyanin of *Helix pomatia*. Eur. J. Biochem. **249**, 195–222
14 Hall, R. L., Kamerling, J. P., Gerwig, G. J. and Vliegenthart, J. F. G. (1977) 3-*O*-methyl sugars as constituents of glycoproteins. Identification of 3-*O*-methyl galactose and 3-*O*-methyl mannose in pulmonate gastropod hemocyanins. Biochem. J. **165**, 173–176
15 Stoeva, S., Rachev, R., Severov, S., Voelter, W. and Genov, N. (1995) Carbohydrate content and monosaccharide composition of *Rapana thomasiana* grosse (Gastropoda) hemocyanin and its structural subunits. Comparison with gastropodan hemocyanins. Comp. Biochem. Physiol. Part B: Biochem Mol. Biol. **110**, 761–765
16 Stoeva, S., Schutz, J., Gebauer, W., Hundsdorfer, T., Manz, C., Markl, J. and Voelter, W. (1999) Primary structure and unusual carbohydrate moiety of functional unit 2-c of keyhole limpet hemocyanin (KLH). Biochim. Biophys. Acta **1435**, 94–109
17 Harris, J. R. and Markl, J. (2000) Keyhole limpet hemocyanin: molecular structure of a potent marine immunoactivator. A review. Eur. Urol. **37**, 24–33
18 Dolashka, P., Genov, N., Parvanova, K., Voelter, W., Geiger, M. and Stoeva, S. (1996) *Rapana thomasiana* grosse (gastropoda) haemocyanin: spectroscopic studies of the structure in solution and the conformational stability of the native protein and its structural subunits. Biochem. J. **315**, 139–144
19 Dolashka-Angelova, P., Schick, M., Stoeva, S. and Voelter, W. (2000) Isolation and partial characterization of the N-terminal functional unit of subunit RtH1 from *Rapana thomasiana* grosse hemocyanin. Int. J. Biochem. Cell Biol. **32**, 529–538
20 Boteva, R., Severov, S., Genov, N., Beltramini, M., Filippi, B., Ricchelli, F., Tallandini, L., Pallhuber, M. M., Tognon, G. and Salvato, B. (1991) Biochemical and functional characterization of *Rapana thomasiana* hemocyanin. Comp. Biochem. Physiol. Part B: Biochem. Mol. Biol. **100**, 493–501
21 Francois, C., Marshall, R. D. and Neuberger, A. (1962) Carbohydrates in protein. Biochem. J. **83**, 335–341
22 Settineri, C. A. and Burlingame, A. L. (1997) Structural characterization of protein glycosylation using HPLC/electrospray ionization mass spectrometry and glycosidase digestion. Methods Mol. Biol. **61**, 254–278
23 Hsi, K.-L., Chen, L., Hawke, D. H., Zieske, L. R. and Yuan, P.-M. (1991) A general approach for characterizing glycosylation sites of glycoproteins. Anal. Biochem. **198**, 238–245
24 Harvey, J. D. (1996) Identification of cleaved oligosaccharides by matrix-assisted laser desorption/ionization. Methods Mol. Biol. **61**, 243–253
25 Harvey, J. D. (1996) Matrix-assisted laser desorption/ionisation mass spectrometry of oligosaccharides and glycoconjugates, J. Chromatogr., A **720**, 429–446
26 Sutton, C. W., O'Neill, J. A. and Cottrell, J. S. (1994) Site-specific characterization of glycoprotein carbohydrates by exoglycosidase digestion and laser desorption mass spectrometry. Anal. Biochem. **218**, 34–46
27 Cuff, M. E., Miller, K. I., van Holde, K. E. and Hendrickson, W. A. J. (1998) Crystal structure of a functional unit from *Octopus* hemocyanin. Mol. Biol. **278**, 855–870
28 Stoeva, S., Idakieva, K., Genov, N. and Voelter W. (1997) Complete amino acid sequence of dioxygen-binding functional unit of the *Rapana thomasiana* hemocyanin. Biochem. Biophys. Res. Commun. **238**, 2, 403–410
29 Gielens, C., De Geest, N., Xin, X.-Q., Devreese, B., van Beeumen, J. and Préaux, G. (1997) Evidence for a cysteine-histidine thioether bridge in functional units of molluscan haemocyanins and location of the disulfide bridges in functional units d and g of the βC-haemocyanin of *Helix pomatia*. Eur. J. Biochem. **248**, 879–888
30 Stoeva, S., Idakieva, K., Betzel, C., Genov, N. and Voelter, W. (2002) Amino acid sequence and glycosylation of functional unit RtH2-e from *Rapana thomasiana* (Gastropod) hemocyanin. Arch. Biochem. Biophys. **399**, 149–158

Received 21 February 2003/9 April 2003; accepted 1 May 2003
Published as BJ Immediate Publication 6 May 2003, DOI 10.1042/BJ20030291