Insect Immunity. The Inducible Antibacterial Peptide Diptericin Carries Two *O*-Glycans Necessary for Biological Activity[†]

Philippe Bulet,[‡,§] Gilbert Hegy,[‡,∥] Jean Lambert,[§] Alain Van Dorsselaer,[∥] Jules A. Hoffmann,[§] and Charles Hetru*[,§]

Institut de Biologie Moléculaire et Cellulaire CNRS (UPR 9022, "Réponse Immunitaire et Développement chez les Insectes"), 15, rue René Descartes, 67084 Strasbourg Cedex, France, and Laboratoire de Spectrométrie de Masse Bio-Organique de l'Université Louis Pasteur (URA 31 CNRS "Chimie Organique des Substances Naturelles"), 1, rue Blaise Pascal, 67008 Strasbourg Cedex, France

Received November 18, 1994; Revised Manuscript Received February 23, 1995[⊗]

ABSTRACT: A bacterial challenge of larvae of the dipteran insect *Phormia terranovae* induces the rapid synthesis of diptericin, an antibacterial polypeptide, previously characterized at the amino acid level and indirectly by cDNA cloning studies. This 82-residue polypeptide consists of an N-terminal proline-rich domain and a central and C-terminal glycine-rich domain. Using liquid chromatography coupled to electrospray ionization-mass spectrometry, we demonstrate here that this molecule is more complex than anticipated and carries two *O*-substitutions on threonine residues, one in the proline-rich domain (residue 10) and one in the glycine-rich domain (residue 54). These substitutions consist of identical trisaccharides: glucose → galactose → *N*-acetylgalactosamine → (threonine). Treatment of diptericin with *O*-glycosidase, which selectively removes the substitutions without altering the polypeptide proper, abolishes the antibacterial activity, indicating that this posttranslational modification is essential for biological activity of the polypeptide. We also show that diptericin is posttranslationally modified by a C-terminal amidation.

Insects respond to bacterial challenge and to injury by the rapid and transient synthesis of a battery of potent antibacterial peptides [reviewed in Boman et al. (1991), Hoffmann et al. (1993), Hultmark (1993), and Cociancich et al. (1994a)]. These are mostly small, cationic peptides with a large spectrum of activity against Gram positive and/or Gram negative bacteria. Some of the induced molecules, however, are larger polypeptides with molecular masses up to 30 kDa. To date, some 50 inducible antibacterial peptides/polypeptides have been totally or partially characterized. Two well-defined families are easily recognized among these molecules: (i) the cecropins, 4-kDa anti-Gram negative and anti-Gram positive peptides, devoid of cysteine, which form two α-helices [reviewed in Boman et al. (1991)], and (ii) the insect defensins (sapecins), 4-kDa anti-Gram positive peptides with six cysteines engaged in three intramolecular disulfide bridges [reviewed in Hoffmann and Hetru (1992); see also Yamada and Natori (1993)]. The other inducible antibacterial peptides can be grouped as proline-rich peptides (2−3 kDa) and glycine-rich peptides (9−30 kDa). Several proline-rich peptides have been characterized over the last years, namely, apidaecin (Casteels et al., 1989), abaecin (Casteels et al., 1990), drosocin (Bulet et al., 1993), and pyrrhocoricin (Cociancich et al., 1994b). Two of these, drosocin and pyrrhocoricin, carry an *O*-glycosylated substitution. The glycine-rich polypeptides form a relatively heterogeneous group of molecules which have in common a higher than normal content of glycine residues (>12%, occasionally up to 22%).

The present study focuses on diptericins, which are 9-kDa anti-Gram negative polypeptides isolated several years ago from the blood of bacteria-challenged larvae of *Phormia terranovae* (Dimarcq et al., 1988). Diptericins have an N-terminal proline-rich domain which extends over 17 residues and shows sequence similarity with the 19-residue proline-rich peptide pyrrhocoricin from the bug *Pyrrhocoris apterus* (Cociancich et al., 1994b). Their central and C-terminal region is rich in glycine residues. The proline-rich domain of diptericin contains a consensus site for *O*-glycosylation (Pro-Thr/Ser-Xaa-Xaa-Pro) (Wilson et al., 1991) around threonine-10. In the study of Dimarcq et al. (1988), the isolated diptericins had not been subjected to mass spectrometry analysis, and posttranslational modifications had not been reported. In view of our recent results with pyrrhocoricin and drosocin, we have now reexamined the structure of diptericin from *Phormia terranovae*. We show that the N-terminal proline-rich domain actually carries a trisaccharide substitution on threonine-10. We also observe a second *O*-glycosylated substitution within the glycine-rich domain, which is not situated within an established consensus sequence for *O*-glycosylation. Interestingly, the removal of the substitutions by *O*-glycosidase treatment abolishes the antibacterial activity, indicating that this posttranslational modification is essential for biological activity.

MATERIALS AND METHODS

Insect Immunization and Hemolymph Collection. Third-instar wandering larvae of *Phormia terranovae* (Diptera) were isolated and kept at 20 °C. They were individually pricked with a fine needle dipped into a combined bacterial pellet obtained after centrifugation of 37 °C overnight

[†] We thank Rhône Poulenc Rorer for financial support of G.H. through the program BioAvenir.

[‡] The first two authors contributed equally to this study.

* To whom correspondance should be addressed.

[§] Institut de Biologie Moléculaire et Cellulaire CNRS.

[∥] Laboratoire de Spectrométrie de Masse BIo-Organique de l'Université Louis Pasteur.

[⊗] Abstract published in *Advance ACS Abstracts,* May 15, 1995.

0006-2960/95/0434-7394$09.00/0

cultures of *Escherichia coli* 1106 (Gram negative bacteria) and *Micrococcus luteus* (Gram positive bacteria). The insects were kept at room temperature for 24 h, and the hemolymph was collected by pricking the anterior part of the larvae, after which the body was gently squeezed to press out a drop (about 15 μL) of hemolymph, which was recovered in a precooled sterile polypropylene tube containing 10 μL of aprotinin as protease inhibitor (final concentration: 10 μg/mL of hemolymph) and phenylthiourea (final concentration: 1 μg/mL of hemolymph) to prevent melanization. The immune hemolymph was rapidly centrifuged (13000*g*, 30 min, 4 °C), and the cell-free supernatant was directly used for the purification of the antibacterial molecules.

Bacteria. Escherichia coli D31 (streptomycin resistant) was a gift from Dr. H. G. Boman (University of Stockholm); *E. coli* D22 was from Dr. P. L. Boquet (Centre d'Etudes Nucléaires, Saclay); *E. coli* 1106 was from Dr. T. Achstetter (Transgène, Strasbourg); and *Micrococcus luteus* A270 was from the Pasteur Institute Collection, Paris. All strains were grown on Luria-Bertani's rich nutrient medium (Bactotrypton 1%, yeast extract 0.5%, NaCl 1%; w/v).

Plate Growth Inhibition Assay. The antibacterial activity was routinely monitored by a plate growth inhibition assay on *E. coli* D22 and *E. coli* D31 as previously described in Lambert et al. (1989).

Purification of the Diptericins from Phormia terranovae. Step I: Sep-Pak Prepurification. The cell-free hemolymph (supernatant, see above) was diluted (v/v) with acidified water (0.1% TFA) and loaded onto Sep-Pak C_{18} cartridges. After washing with 5 mL of acidified water (0.05% TFA), elutions were performed with solutions of 15%, 40%, and 80% acetonitrile in acidified water (0.05% TFA). All fractions were concentrated in a vacuum centrifuge (Savant) to remove the organic solvent and TFA and were reconstituted with MilliQ water. The presence of antibacterial activity was detected with the plate growth inhibition assay.

Step II: Size-Exclusion Chromatography. The first step of purification was a size fractionation by a high-performance gel permeation chromatography system (HPGPC) consisting of serially linked SEC 3000 and SEC 2000 columns (300 × 7.5 mm, Beckman). As the bulk of antibacterial activity during the extraction with Sep-Pak C_{18} cartridges was found to be present in the 40% elution fraction (see above), only this fraction was applied on the columns and eluted in isocratic conditions with 30% acetonitrile in acidified water (0.05% TFA) at a flow rate of 0.5 mL/min. The fractions eluted from the columns were subjected to ultraviolet absorbance at 225 nm and to the plate growth inhibition assay.

Step III: Reversed-Phase HPLC. The fractions that contained the antibacterial activity were applied to an Aquapore OD 300 C_{18} column (220 × 4.6 mm, Brownlee Associates) equilibrated with 10% acetonitrile in acidified water. Elution was performed with a linear gradient of 10−60% acetonitrile in acidified water over 120 min at a flow rate of 1 mL/min. The column effluent was followed by absorbance at 225 nm, and the presence of antibacterial activity was monitored as above.

Step IV: Final Purification. The fractions that contained the antibacterial activity were applied to an Aquapore OD 300 C_{18} column (220 × 4.6 mm, Brownlee Associates). The column was developed with a linear biphasic gradient from 2 to 21% acetonitrile in acidified water over 10 min and then from 21 to 33% acetonitrile in acidified water over 60 min at a flow rate of 0.8 mL/min. Antibacterial activity was detected as above.

All HPLC purifications were performed with a Beckman Gold HPLC system equipped with a Beckman 168 photodiode array detector.

Microsequence Analysis. Automated Edman degradation of peptides in their native form and after enzymatic digestion and detection of phenylthiohydantoin derivatives was performed on a pulse liquid automatic sequenator (Applied Biosystems, Model 473 A).

Enzymatic Digestions and Separation of Products of Digestion. (A) Arginyl Endopeptidase Digestion. Digestion of 0.5 nmol of pure peptide with arginyl endopeptidase (Arg-C, Takara, Kyoto) was performed in 10 mM Tris-HCl buffer, pH 8, in the presence of 0.01% Tween 20, during 15 h at 37 °C with a peptide/enzyme ratio of 100:1 (w/w). The digestion was stopped by adding 0.1% TFA; separation and collection of peptide fragments were performed by HPLC.

(B) Tryptic Digestion. One nanomole of pure peptide was treated with trypsin (Boehringer Mannheim) at a peptide/enzyme ratio of 100:1 (w/w) for 15 h at 37 °C in 100 mM Tris-HCl buffer, pH 8.5. The digestion was stopped as previously, and fragments were analyzed by liquid chromatography−electrospray ionization−mass spectrometry (LC−ESI−MS).

Mass Spectrometry. The purified peptide was dissolved in water/methanol (50/50, v/v) containing 1% acetic acid and analyzed on a VG Biotech BioQ mass spectrometer (Manchester). This instrument consists of an electrostatic ion spray source operating at atmospheric pressure, followed by a quadrupole mass analyzer with a mass range of 1−4000. The extraction cone voltage value was 55 V. Scanning was performed from $m/z = 500$ to 1500 in 10 s with the resolution adjusted so that the $m/z = 998$ peak from horse heart myoglobin was 1.5−1.7 Da wide at its base. The data system was operated as a multichannel analyzer, and several scans were summed to obtain the final spectrum. Each molecular species produced a series of multiply charged protonated molecular ions from which the molecular mass was determined by simple calculation. Calibration was performed using the multiply charged ions from a separate introduction of horse heart myoglobin (16 951.4 Da). Molecular masses are given as average values based on the atomic weights of the elements (C = 12.011, H = 1.00794, N = 14.0067, O = 15.9994 and S = 32.06); only average masses were measured.

Identification of Carbohydrates. One hundred microliters of Milli Q water was added to 1 nmol of glycopeptide. The solubilized material was transferred into a glass vial and dried in a vacuum centrifuge (Savant). One hundred microliters of a solution of 1 N HCl in methanol was added to the dried glycopeptide, which was maintained in a hermetically closed vial and incubated at 70 °C for 9 h. HCl/methanol (1 N) was prepared by diluting a HCl/methanol (3 N) solution (Supelco) with anhydrous methanol under an anhydrous N_2 atmosphere. Anhydrous methanol was obtained by distillation after refluxing with CaH_2 for 30 min. The material was dried again in a vacuum centrifuge and derivatized with 10 μL of *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide in anhydrous pyridine and trimethylchlorosilane (1/1/0.01, v/v/v). Analysis was made by gas chromatography−mass

spectrometry (GC−MS) on a Fisons MD800 mass spectrometer. The gas chromatography (GC) equipment included a fused silica column (DB5; 30 m; 0.1 μm film thickness; 0.32 mm i.d.) and a moving needle ROS injector heated at 250 °C and was programmed from 80 to 200 °C at a rate of 4 °C/min. Helium was used as carrier gas (1.2 mL/min) and 2−50 ng of material (including the silylating reactives) were injected. The mass spectrometer was operated in electron impact (EI) mode. Mass spectrometric conditions were as follows: electron energy, 70 eV; trap current, 150 μA; mass range scanned from *m/z* 50 to 650; scan speed, 1 s.

The components were identified by comparison with retention times and spectra of authentic sugars submitted to the same treatment as the native glycopeptide.

Liquid Chromatography−Electrospray Ionization−Mass Spectrometry (LC−ESI−MS). LC−ESI−MS analyses of peptide mixtures obtained by tryptic digestion were carried out using a 140A syringe pump (Applied Biosystems) coupled to a VG Biotech BioQ mass spectrometer (Manchester) operated in the electrospray pneumatically assisted ionization mode. The setup for interfacing the ion spray source to the HPLC has been described elsewhere (Klarskov et al., 1994). RP-HPLC conditions involved gradient elution using two mobile phases. Solvent A consisted of acidified water (0.1% TFA), and solvent B was acetonitrile containing 0.08% TFA. The gradient increased from 2 to 52% B in 50 min and from 52 to 80% B in 5 min and was followed by isocratic elution at 80% B during 5 min. Peptides were separated on a reversed-phase column (Macherey-Nagel, C_{18}, 5 μm, 2.1 × 125 mm, Düren, Germany) with a flow rate of 0.2 mL/min and were detected at 214 nm by a Waters 486 tunable absorbance detector. The column effluent was divided by a Valco tee (Valco, Houston) between mass spectrometer and UV detector with a split of 1/5. An additional liquid sheath containing 2-methoxyethanol/2-propanol (2/1, v/v) and 1% formic acid was mixed at the tip top. The mass spectrometer was scanned over a range of *m/z* = 400−2000 at 6 s/scan. Calibration was performed using a mixture of polyethylene glycol with a molecular weight of 400 Da and 2000 Da.

O-Glycosidase Treatment. Diptericins were digested with 0.1 munit of *O*-glycan-peptidehydrolase (*O*-glycosidase, E.C. 3-2-1-97, Boehringer Mannheim) according to the supplier's instruction. Briefly the polypeptide (0.1 μg) was incubated with the enzyme in 100 μL of 20 mM, sodium cacodylate buffer, pH 6, for 15 min at 37 °C. The enzyme reaction was stopped by the addition of TFA (final concentration, 0.05%), and after evaporation and uptake in Milli Q water (4 μL), the antibacterial activity was tested by the plate growth inhibition assay. Three different controls were performed: (i) diptericins were incubated with the medium in the absence of enzyme, (ii) medium was incubated with enzyme in the absence of diptericins, and (iii) diptericins to which had been added the reaction mixture previously stopped by TFA.

RESULTS

Larvae of *Phormia terranovae* (550) were challenged by injection of a low dose of bacteria, and after 24 h their hemolymph (8 mL) was collected, freed from hemocytes by centrifugation, and applied to six serially linked Sep-Pak C_{18} cartridges for solid-phase extraction. Aliquots of the fractions were tested by the plate growth inhibition assay against *E. coli* (strains D22 and D31). A strong anti-Gram negative activity was recovered by elution with 40% acetonitrile in acidified water (0.05% TFA). This active material was applied to high-performance gel permeation chromatography columns developed in isocratic conditions with 30% acetonitrile in acidified water (0.05% TFA). Anti-Gram negative activity was detected in one large absorption peak corresponding to compounds with molecular masses ranging from 5 to 15 kDa.

The active material was subjected to C_{18} reversed-phase HPLC and eluted with a gradient of acetonitrile in acidified water (Materials and Methods, Step III). Five anti-Gram negative compounds with closely related chromatographic behaviors were recovered (Figure 1) and referred to as compounds α to ε. They were further individually purified to apparent homogeneity by HPLC (Materials and Methods, Step IV). The pure products were submitted to Edman degradation, and for the predominant molecule (γ, strongest UV absorption) the sequence of 64 N-terminal residues was directly obtained (Figure 2) with two ambiguities in positions 10 and 54. This amino acid sequence is identical to that deduced from cDNA cloning studies of *Phormia* diptericin (Reichhart et al., 1989) where the amino acids in positions 10 and 54 are encoded by threonine codons. The other four anti-Gram negative molecules, α, β, δ, and ε (Figure 1), were also directly sequenced. Only partial N-terminal sequences were obtained (varying from 20 to 57 residues), which were all in agreement with that of compound γ. They all exhibited an ambiguous signal for residue 10, and in one of the four isoforms (β) for which the sequence extended to residue 57, threonine at position 54 gave also an ambiguous signal.

The two predominant isoforms (β, γ) were subjected to enzymatic cleavage with arginyl endopeptidase, and the fragments were analyzed by Edman degradation after HPLC purification. The complete peptide sequence of both isoforms was identical with ambiguous signals for residues 10 and 54 (Figure 2).

The diptericin isoforms α to ε were next submitted to mass spectrometry and yielded distinct molecular masses. The predominant molecule γ gave a molecular mass of 9423.3 Da, whereas the masses of the four other molecules were, in increasing order, as follows: 8895.8 Da (ε), 9099.6 Da (δ), 9260.6 Da (β), and 9745.7 Da (α). The differences between the observed molecular masses and the calculated masses (assuming threonine residues in 10 and 54), are multiples of 203 and/or 162 Da (Table 1). These differences could correspond to an *N*-acetylhexosamine (221, i.e., 203 + 18) and a hexose (180, i.e., 162 + 18), assuming the loss of one molecule of water in the formation of *O*-glycosylated bonds. As anhydrous hydrofluoric acid (AHF) treatment selectively cleaves *O*-glycosylated bonds, resulting in the deglycosylation of the peptide backbone, we subjected native diptericin γ to this treatment and observed that its mass was reduced from 9423.3 to 8692.6 Da, which is in perfect agreement with the calculated mass of the peptide sequence with threonine residues in positions 10 and 54.

To identify the sugar moieties linked to the threonine residues, we have subjected 1 nmol of the γ isoform (9423.3 Da) to HCl/methanol treatment, which cleaves carbohydrates from the peptide backbone. The resulting mixture was pertrimethylsilylated and analyzed by GC−MS. Two car-

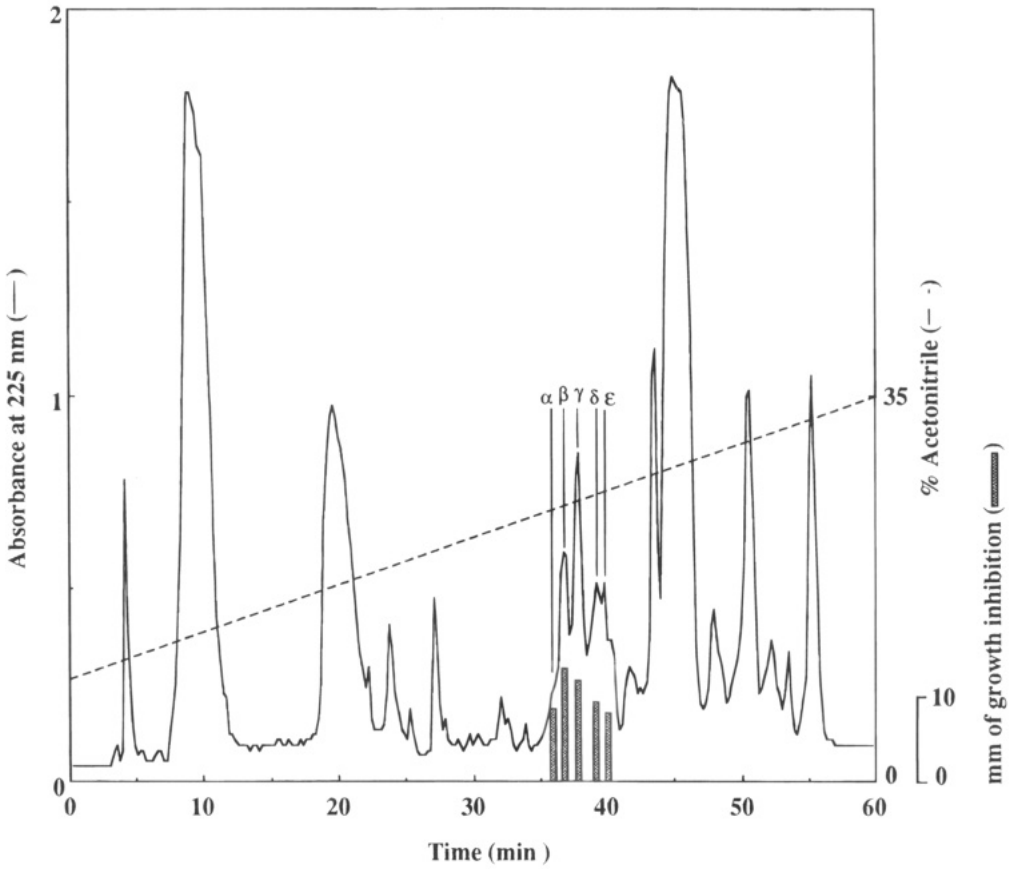

FIGURE 1: Reversed-phase HPLC separation of immune hemolymph of *Phormia terranovae*. Elution was performed over 120 min with a linear gradient (dashed line) of 10−60% acetonitrile in acidified water (0.05% TFA) at a flow rate of 1 mL/min. Absorbance was monitored at 225 nm (solid line). The antibacterial activity against Gram negative bacteria was tested on aliquots of each fraction by the plate growth inhibition assay and is expressed in diameter of growth-inhibition zone (shaded columns). Only the chromatogram zone containing diptericin-like material is presented (from 0 to 60 min). The denominations of the active peaks, α to ϵ, are given in relation with the retention time.

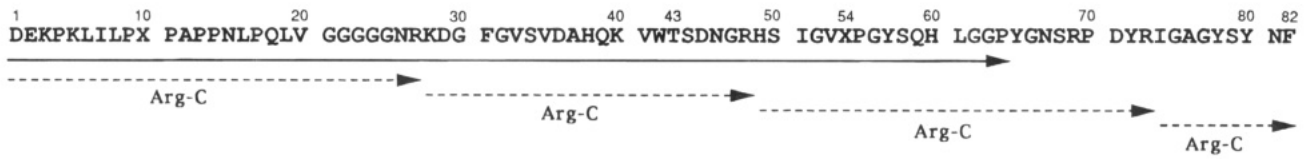

FIGURE 2: Amino acid sequence of diptericin γ determined by Edman degradation. The N-terminal amino acid sequence obtained by Edman degradation of diptericin (solid arrow) and sequences of fragments obtained after arginyl endopeptidase cleavage (dashed arrows, Arg-C) are shown assembled into one complete overlapping sequence of 82 residues. The two Xs (residues 10 and 54) in the sequence indicate the positions of ambiguous signals observed during the sequencing process.

Table 1: Molecular Masses of the Different Diptericin Isoforms Isolated in Extracts of *Phormia terranovae* Larvae (α to ϵ)[a]

diptericin isoforms	molecular masses of native diptericins (Da)	mass differences compared to calculated mass (Da)	possible mass interpretations of the differences observed (Da)
ϵ	8895.8	203.3	203
δ	9099.6	407.0	2 × 203 (406)
β	9260.6	568.0	2 × 203 + 162 (568)
γ	9423.3	730.7	2 × 203 + 2 × 162 (730)
α	9745.7	1053.1	2 × 203 + 4 × 162 (1054)

[a] The differences with the molecular mass calculated for diptericin (8692.6) deduced from cDNA studies, have been reported and these differences have been speculated to be multiples of the masses of hexoses and N-acetylhexosamines.

bohydrates were detected, which showed retention times (Figure 3) and fragmentation spectra (data not shown) respectively identical to those of galactose and *N*-acetylgalactosamine. Smaller amounts of the 9745.7-Da (α) molecule were submitted to a similar analysis and yielded, in addition to *N*-acetylgalactosamine and galactose, another signal which had the same retention time (Figure 3) and the same fragmentation spectrum as glucose.

To analyze separately the substitution on each of the threonines, 1 nmol of the γ isoform (9423.3 Da) was subjected to enzymatic cleavage with trypsin followed by liquid chromatography−electrospray ionization−mass spectrometry (LC−ESI−MS) analysis, which yielded eight ion current peaks (Figure 4). Nine distinct mass fragments were analyzed, one for each current peak with the exception of peak 8, which contained two fragments. For each fragment, the molecular mass was determined and the amino acid sequence was established by Edman degradation (Table 2, numbering of amino acids as in the full diptericin sequence presented in Figure 2). The molecular mass analysis indicated that four fragments carried a substitution: of these fragments one present in peak 5, one in peak 6, and two in

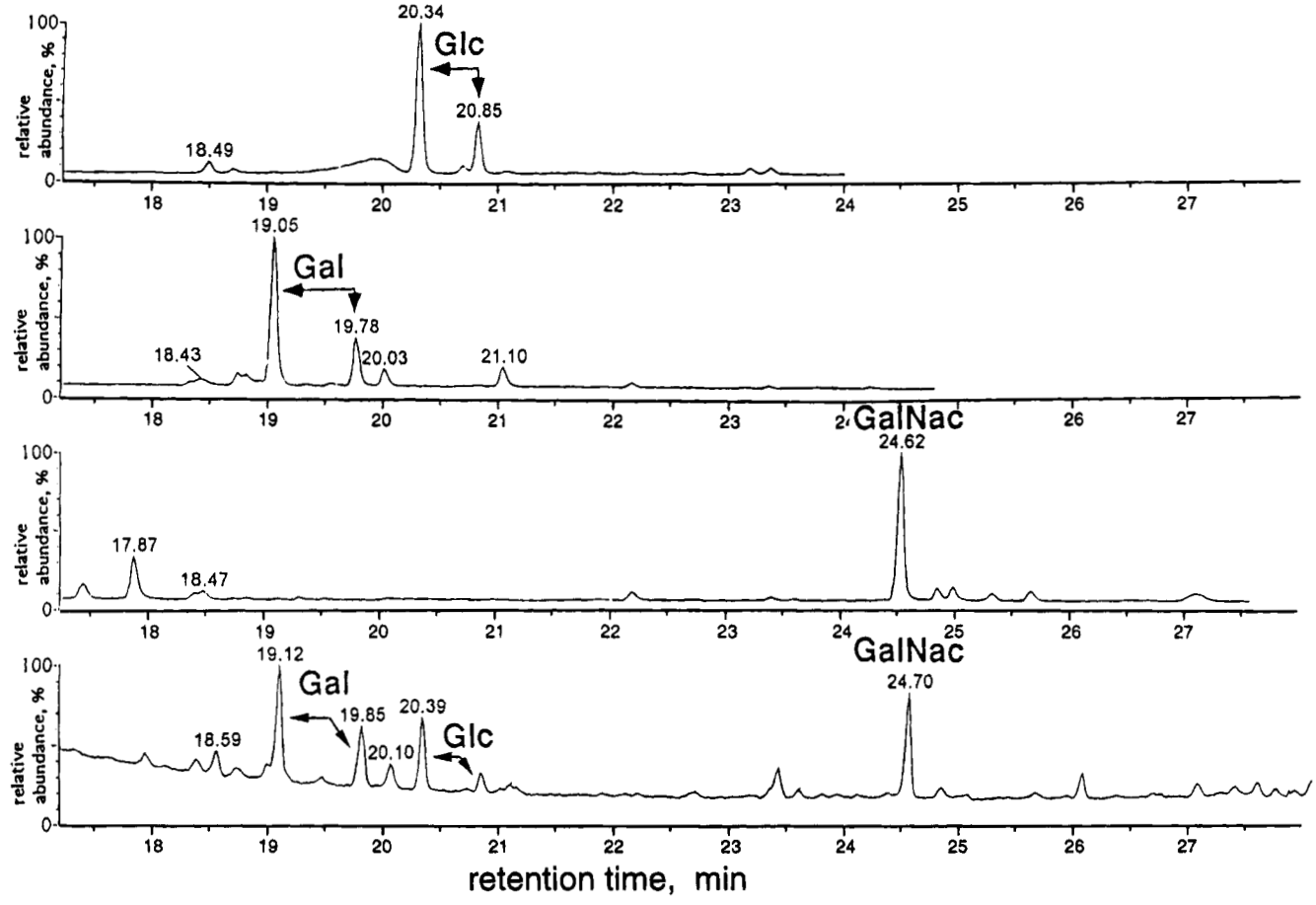

FIGURE 3: Identification of carbohydrate composition by comparison with standards. Gas chromatography−mass spectrometry analysis of the standards glucose (Glc), galactose (Gal), and *N*-acetylgalactosamine (GalNac) is shown by the three upper traces. The native glycopeptide submitted to methanolysis and pertrimethylsilylation yielded peaks in the chromatographic zone corresponding to carbohydrate with retention times of glucose, galactose, and *N*-acetylgalactosamine. The number at the top of each peak indicates the retention time. The *x*-axis represent the number of mass spectra acquired during the chromatographic analysis (one scan corresponds to 3 s).

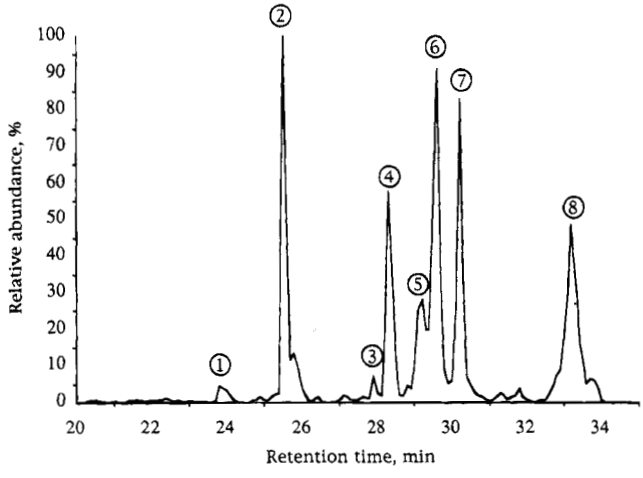

FIGURE 4: On-line LC−ESI−MS analysis of tryptic digest of the antibacterial peptide (diptericin γ). One nanomole of the glycopeptide was analyzed by LC−ESI−MS on a narrow-bore column with a postcolumn split. The *y*-axis represents the peak intensity plot of the measured ion current. The *x*-axis shows the retention time of the mass spectrum acquired during the chromatographic analysis. The peptide sequence corresponding to each peak is given in Table 2 in relation with the number indicated.

peak 8 (Figure 4; Table 2). The two fragments in peak 8 differed only by the presence or absence of a C-terminal lysine, and each contained the substituted threonine in position 10 with an excess of mass corresponding to *N*-acetylgalactosamine and galactose. The fragments from peaks 6 and 5 had the same amino acid sequence as the corresponding stretch of residues in diptericin, but differed by the presence of an excess of mass of respectively 203 and 365, from the diptericin sequence. These differences correspond to the substitution by an *N*-acetylgalactosamine, and a galactose → *N*-acetylgalactosamine disaccharide.

We conclude from these data that in peptide γ, the substitution on threonine-54 corresponds to the disaccharide galactose → *N*-acetylgalactosamine. We also propose that the same disaccharide is carried by threonine-10 in this molecule (see Discussion).

The analysis of the low molecular weight fragment in peak 7 indicates that the C-terminus of diptericin is amidated. Indeed, the amino acid sequence established by Edman degradation corresponds to that of the nine C-terminal residues (Table 2), but the experimental mass of 990.6 Da determined by LC−ESI−MS is lower by 1 Da than the calculated mass (991.6 Da). This difference is best explained by assuming that the C-terminal phenylalanine residue is amidated. This inference is in agreement with the presence of a glycine codon at the C-terminus in the diptericin cDNA (Reichhart et al., 1989).

The fragment detected in peak 2 contained the threonine residue of position 43 of diptericin. Its molecular mass clearly showed that this fragment does not carry any substitution.

Table 2: Analysis of Peaks Observed in LC–EIS–MS of Trypsin Fragments of Diptericin γ[a]

peak no.	retention time	exptl molecular masses	calcd molecular masses	peptidic sequences with deduced posttranslational modifications
1	23.9	615.7	615.7	1 5 DEKPK
2	25.5	933.6	933.6	41 48 VWTSDNGR
3	27.9	1387	1387	28 40 KDGFGVSVDAHQK
4	28.3	1289	1289	29 40 DGFGVSVDAHQK
5	29.2	3081	2715	49 73 HSIGVTPGYSQHLGGPYGNSRPDYR Gal->GalNac
6	29.6	2919	2715	49 73 HSIGVTPGYSQHLGGPYGNSRPDYR GalNac
7	30.2	990.6	991.6	74 82 IGAGYSYNF-NH_2
8	33.2	2631 and 2504	2266 and 2139	6 28 LILPTPAPPNLPQLVGGGGGNRK Gal->GalNac 6 27 LILPTPAPPNLPQLVGGGGGNR Gal->GalNac

[a] Column 1 indicates peak numbers in the chromatogram presented in Figure 4. Column 2 gives the retention time (min) for each peak. Column 3 shows the observed molecular masses for each peak. Column 4 gives the molecular masses calculated for the corresponding unsubstituted amino acid sequences. Column 5 presents the peptide sequences and the glycosylations inferred from the molecular masses observed. Four of the fragments can only be explained by the presence of *O*-glycosylated substitutions.

Finally, we have treated with *O*-glycosidase isoform δ, which carries two *N*-acetylgalactosamines, and isoform γ, which carries two galactose → *N*-acetylgalactosamine disaccharides. This enzyme specifically removes the disaccharide galactose → *N*-acetylgalactosamine but does not affect the monosaccharide *N*-acetylgalactosamine. After treatment (Materials and Methods), the diptericins were tested by the plate growth inhibition assay, and their activity was compared to that of untreated diptericins δ and γ, which served as controls. Diptericin γ completely lost its activity upon enzymatic cleavage, while diptericin δ treated with *O*-glycosidase had an activity identical to that of the controls. The medium with the enzyme had no antibacterial activity, and the treatment of diptericin without active enzyme did not inhibit the antibacterial activity of the diptericins, a result which is in keeping with the conserved activity of diptericin δ treated by *O*-glycosidase (see above). These data confirm the specificity of the *O*-glycosidase cleavage of the disaccharide galactose → *N*-acetylgalactosamine and show that *O*-glycosidase treatment of diptericin abolishes the antibacterial activity. Given that the presence of a single *N*-acetylgalactosamine in diptericin ϵ confers some antibacterial activity to the polypeptide, we can infer that the treatment with *O*-glycosidase has led to the complete deglycosylation of diptericin. In conclusion, the presence of *O*-glycosylated substitutions is necessary for the biological activity of the diptericin from *Phormia terranovae*.

DISCUSSION

We report in this study the isolation of five diptericin isofoms with close chromatographical behaviors. Edman degradation demonstrates that these isoforms are identical at the level of the amino acid sequence. Our data indicate that the five isoforms carry *O*-glycosylated substitutions of various complexities. We propose that in all the isoforms the substitutions are present on two out of the three threonine residues, i.e., on threonine-10, which is located in an *O*-glycosylation consensus sequence, and threonine-54, the location of which does not correspond to such a consensus sequence. Threonine-43 is not substituted. We also propose that the isoform with the highest molecular mass carries two identical trisaccharide moieties, glucose → galactose → *N*-acetylgalactosamine. These proposals are based on the following results:

(i) Residues 10 and 54, which both correspond to threonine according to the previous cDNA studies (Reichhart et al., 1989), gave ambiguous signals in Edman degradation, which is compatible with the assumption that these residues carry a substitution. In contrast, threonine in position 43, which is not substituted, gave a clear signal in Edman degradation, indicating that this residue is easily identified in our conditions.

(ii) The mass spectrometry analysis of the five isoforms yielded masses which all differed from that calculated for diptericin by multiples of 203 and 162 Da. Taking into account the loss of a molecule of H_2O in the formation of a bond, these are in fact multiples of 221 and 180 Da, which correspond to the masses of *N*-acetylhexosamine and hexose.

(iii) Treatment with anhydrous hydrofluoric acid, which specifically cleaves *O*-saccharide bonds on peptides without interfering with the peptide chain, reduced the mass of the major diptericin from 9423.3 Da (γ isoform) to 8692.6 Da, i.e., a mass identical to that calculated from the peptide sequence.

(iv) Treatment of diptericin with HCl in anhydrous methanol, which cleaves the sugar bonds, followed by pertrimethylsilylation, yielded volatile sugar derivatives identified by gas chromatography–mass spectrometry as

N-acetylgalactosamine, galactose, and glucose.

(v) When diptericin isoforms carrying three, four, or six sugar moities were submitted to additional purification steps, we systematically observed their transformation to less substituted isoforms. This observation is explained by the sensivity of the glycosidic substitutions to the acidic purification conditions. In contrast, the isoform carrying only *N*-acetylgalactosamines was not affected by the purification conditions, probably because the bond between the amino acid and the sugar is not an osidic bond. We interpret the presence of various isoforms of diptericins observed in our extracts after purification as the result of limited degradation of one single complex molecule which carries two trisaccharide substitutions (isoform α).

The analysis of the tryptic fragments of diptericin γ by liquid chromatography−mass spectrometry shows that the substitution on threonine-54 is a galactose → *N*-acetylgalactosamine disaccharide. The substitution on threonine-10 in this isoform can be identified as follows: isoform δ carries two *N*-acetylgalactosamine substitutions, and as only one *N*-acetylgalactosamine is linked to threonine-54, the second saccharide is by necessity linked to threonine-10. In fact the molecular mass of the fragment present in peak 8 (Table 2) indicates that threonine-10 is substituted by the dissaccharide galactose → *N*-acetylgalactosamine. Finally, in the case of isoform α, the glucose is obviously in the most distal position in the trisaccharide glucose → galactose → *N*-acetylgalactosamine.

It is an open possibility that the *O*-substitutions carried by diptericin are more complex. Our extraction procedures involve acidic conditions which could cleave some sugar bonds and, namely, cleave off sialic acids which often terminate oligosaccharides. The number of known structures of carbohydrate units of glycoproteins has grown immensely, and the early assumption that living organisms form only an exceedingly small fraction of the theoretically possible molecular permutations of the dozen or so monosaccharides typically found in glycoconjugates seems no longer justified (Lis & Sharon, 1993). In insects, which present an extreme diversity in morphology and physiology, the information on carbohydrates in glycoproteins is until now very reduced.

It is of interest to note that the biological activity of diptericin is lost when the *O*-substitution is removed. At present the mode of action of diptericin is unknown, primarily because the native material isolated so far was insufficient for detailed studies on the mode of action. It is therefore difficult to propose a role for the substitution in the mode of action of the molecule. For most glycoproteins the role of the carbohydrate moiety is still obscure. For *O*-linked sugars, there is evidence for participation in biological phenomena such as cell−cell interactions [reviewed in Paulson (1989) and Brandley and Schnaar (1986)], but precise biological roles have only been shown for a few glycoproteins. One apparently universal consequence of *O*-glycosylation is the relative resistance to proteases of the *O*-glycosylated regions in the glycoproteins. The most likely explanation for protease resistance is that the attached carbohydrate blocks access to the peptide core, since these same peptide sequences become susceptible to proteases in the absence of the attached carbohydrate. A second consequence of *O*-glycosylation is the induction of a specific conformation. Thus, in the case of diptericin, we can speculate that the molecule interacts with a receptor which accommodates only the glycosylated form in the active conformation. Glycosylated diptericin might also exert a competitive inhibition of the enzymes of bacterial peptidoglycan biosynthesis. We will address these questions as sufficient biological material becomes available.

ACKNOWLEDGMENT

We are indebted to Gérard Teller for mass spectrometry analysis of the carbohydrates.

REFERENCES

Boman, H. G., Faye, I., Gudmundsson, G. H., Lee, J. Y., & Lidholm, D. A. (1991) *Eur. J. Biochem. 201*, 23−31.

Brandley, B. K., & Schnaar, R. L. (1986) *J. Leuk. Biol. 40*, 97−111.

Bulet, P., Dimarcq, J. L., Hetru, C., Lagueux, M., Charlet, M., Hegy, G., Van Dorsselaer, A., & Hoffmann, J. A. (1993) *J. Biol. Chem. 268*, 14893−14897.

Casteels, P., Ampe, C., Jacobs, F., Vaeck, M., & Tempst, P. (1989) *EMBO J. 8*, 2387−2391.

Casteels, P., Ampe, C., Riviere, L., Van Damme, J., Elicone, C., Fleming, M., Jacobs, F., & Tempst, P. (1990) *Eur. J. Biochem. 187*, 381−386.

Cociancich, S., Bulet, P., Hetru, C., & Hoffmann, J. A. (1994a) *Parasitol. Today 10*, 132−139.

Cociancich, S., Dupont, A., Hegy, G., Lanot, R., Holder, F., Hetru, C., Hoffmann, J. A., & Bulet, P. (1994b) *Biochem. J. 300*, 567−575.

Dimarcq, J. L., Keppi, E., Dunbar, B., Lambert, J., Reichhart, J. M., Hoffmann, D., Rankine, S. M., Fothergill, J. E., & Hoffmann, J. A. (1988) *Eur. J. Biochem. 171*, 17−22.

Hoffmann, J. A., & Hetru, C. (1992) *Immunol. Today 13*, 411−415.

Hoffmann, J. A., Hetru, C., & Reichhart, J. M. (1993) *FEBS Lett. 325*, 63−66.

Hultmark, D. (1993) *Trends Genet. 9*, 178−183.

Klarskov, K., Roecklin, D., Bouchon, B., Sabatié, J., Van Dorsselaer, A., & Bischoff, R. (1994) *Anal. Biochem. 216*, 127−134.

Lambert, J., Keppi, E., Dimarcq, J. L., Wicker, C., Reichhart, J. M., Dunbar, B., Lepage, P., Van Dorsselaer, A., Hoffmann, J. A., Fothergill, J., & Hoffmann, D. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 262−266.

Lis, H., & Sharon, N. (1993) *Eur. J. Biochem. 218*, 1−27.

Paulson, J. C. (1989) *Trends Biochem. Sci. 14*, 272−275.

Reichhart, J. M., Essrich, M., Dimarcq, J. L., Hoffmann, D., Hoffmann, J. A., & Lagueux, M. (1989) *Eur. J. Biochem. 182*, 423−427.

Wilson, I. B. H., Gavel, Y., & Von Heijne, G. (1991) *Biochem. J. 275*, 529−534.

Yamada, K., & Natori, S. (1993) *Biochem. J. 291*, 275−279.

BI942678V