Terminal disialylated multiantennary complex-type N-glycans carried on acutobin define the glycosylation characteristics of the Deinagkistrodon acutus venom

Chia-Wei Lin², Jin-Mei Chen³, Ying-Ming Wang³, Sz-Wei Wu³, Inn-Ho Tsai¹,²,³, and Kay-Hooi Khoo¹,²,³

¹Institute of Biochemical Sciences, National Taiwan University, Taipei 106, Taiwan and ²Institute of Biological Chemistry, Academia Sinica, Taipei 115, Taiwan

Glycosylation analysis of nonmammalian sources often springs surprises and conjures up intriguing views of evolutionary adaptation. Many of the constituents of snake venom are known to be glycosylated and yet very few were fully characterized and accorded specific functions. In the process of glycomic screening through the venoms from Asian pit vipers, a partially O-acetylated NeuAcα2-8 NeuAcα2-3 Galβ1-4 GlcNAcβ1-terminal epitope was found to be the predominant glycosylation characteristic of the snake venom produced by the monotypic Deinagkistrodon acutus, with acutobin, a highly specific fibrinogenase, being identified as a primary protein carrier. Full structural definition and glycosylation site mapping were completed through advanced mass spectrometry analyses at both the glycan and glycopeptide levels in conjunction with chemical and enzymatic cleavages. Although similar occurrence of such terminal disialyl cap on the N-glycans of several mammalian glycoproteins has been implicated, most of these correspond to only minor constituents of the full glycomic heterogeneity and remain poorly characterized. In contrast, each antennae of the hybrid- and complex-type N-glycans derived from acutobin was found to be rather homogeneously disialylated. With up to eight sialic acids evenly distributed on nonextended tetraantennary core structure, these unusual N-glycans are among those of highest sialic acid density ever identified without actually carrying polysialic acid chains. It remains to be tested whether they may serve as multivalent disialyl ligands for several of the human Siglecs and thus meddle with the natural immuno-recognition systems of snakebite victims, apart from affecting the general efficacy of acutobin as anticoagulant in biomedical applications.

Keywords: acutobin / disialyl / glycosylation analysis / mass spectrometry / snake venom

Introduction

Snake venom is a complex cocktail of protein families with considerable medical and pharmacological importance. As a natural biological arsenal, the venom is heritable and subjects to accelerated evolution (Kordis and Gubensek 2000; Nikandrov et al. 2005) in adapting to ecological changes. Recent technical advances have facilitated the proteomic characterization of a wide range of the snake venom proteomes (Birrell et al. 2007; Fox and Serrano 2008; Gutierrez et al. 2009). However, despite glycosylation being widely implicated as a common modification on numerous venom proteins and impacts their in vivo venomic functions (Nawarak et al. 2004; Birrell et al. 2006; and reviewed in Soares and Oliveira 2009), systematic glycomic survey has not been undertaken. In fact, very little is known about the glycobiology and range of glycosylation structures that can be synthesized by the snake venom glands in general, with only a handful of the venom glycoproteins from a limited range of species have been investigated for their detailed glycan structures.

The better studied venomous snakes generally fall within the Elapidae and Viperidae families. Among the elapid snake venom glycosylations, the Asian cobra venom glycoproteins were shown to carry terminal α-galactosyl residue (Gowda and Davidson 1992) or, more specifically, in the case of Naja naja kaouthia venom, α3-galactosylated Lewis X (LeX) on the complex-type N-glycans (Gowda et al. 1992) and the O-glycans with poly-N-acetyllactosamine extension (Gowda and Davidson 1994), instead of the more common sialylated termini. Yet, a factor Xa homolog from the venom of an Australian elapid, Tropidechis carinatus, was found to carry sialylated N-glycans and O-Xyl-Glc/GlcNAc (Joseph et al. 2003), thus suggesting significant structural variations, at either intergeneric or interspecies level. Within the viperid snake family, the N-glycans of factor X activator from Russell’s viper venom have been recently analyzed in detail, which substantiated an earlier study reporting the occurrence of α2-3 Neu5Ac sialylated, bisected complex-type N-glycans (Gowda et al. 1994), but additionally identified the prominent presence of terminal sialyl LeX epitope (Chen et al. 2008).
Neither this nor the terminal α3-galactosylated Le^X has been identified in the venoms from the pit vipers. The reported structural details from this subfamily have so far been restricted to the N-glycans of thrombin-like protease ancrod from the monotypic Malayan pit viper (Pfeiffer et al. 1992) and those of batroxobin from a South American pit viper, Bothrops moojeni (Tanaka et al. 1992; Lochnit and Geyer 1995). Interestingly, lacdiNAc, or GalNAcβ1→4GlcNAcβ1-, which can be additionally α2-3Neu5Ac sialylated and fucosylated, was found to be a common structural motif.

From these limited studies, the emerging picture suggests that snake venom glycosylation does share the same basic structural architecture with mammalian N- and O-glycans. However, the sampling pool is still too narrow in phylogenetic distribution to allow a better assessment of whether there are distinctive glycosylation characteristics associated with taxonomical classification of the venomous snakes. We have recently initiated a systematic glycomic survey of the Asian pit viper venoms and identified glycosylation patterns covering a full range of known vertebrate glycan structures, some with unique terminal epitopes. Among these, a rather uncommon disialylated terminal epitope carried on multi-antennary complex-type N-glycans was found to be the defining feature of the glycomic profile of the venom from the monotypic Deinagkistrodon acutus. Remarkably, the occurrence of the disialylated epitope was not detected among the serum proteins from this species, nor in any other Asian pit viper snake venoms examined to date, thus marking it venomic- and potentially species-specific. Following this lead finding, we report here the identification of acutobin, a highly specific fibrinogenase purified from the D. acutus venom, as a major glycoprotein carrier of this unique disialylated epitope, which is now fully characterized at both the glycan and glycopeptide levels by advanced mass spectrometry analyses.

Results

Among the N-glycomic profiles of crude snake venoms from Asian pit vipers, that of the monotypic D. acutus was particularly striking as it carried only a few dominant peaks, with unusually high degree of sialylation in the form of terminal disialylated LacNAc structure (data not shown but similar to Figure 1B). A corresponding shotgun proteomic analysis of disialylated LacNAc sialylated and fucosylated, was found to be a common structural motif.

From these limited studies, the emerging picture suggests that snake venom glycosylation does share the same basic structural architecture with mammalian N- and O-glycans. However, the sampling pool is still too narrow in phylogenetic distribution to allow a better assessment of whether there are distinctive glycosylation characteristics associated with taxonomical classification of the venomous snakes. We have recently initiated a systematic glycomic survey of the Asian pit viper venoms and identified glycosylation patterns covering a full range of known vertebrate glycan structures, some with unique terminal epitopes. Among these, a rather uncommon disialylated terminal epitope carried on multi-antennary complex-type N-glycans was found to be the defining feature of the glycomic profile of the venom from the monotypic Deinagkistrodon acutus. Remarkably, the occurrence of the disialylated epitope was not detected among the serum proteins from this species, nor in any other Asian pit viper snake venoms examined to date, thus marking it venomic- and potentially species-specific. Following this lead finding, we report here the identification of acutobin, a highly specific fibrinogenase purified from the D. acutus venom, as a major glycoprotein carrier of this unique disialylated epitope, which is now fully characterized at both the glycan and glycopeptide levels by advanced mass spectrometry analyses.

**Terminal NeuAc2-8NeuAc disialyl LacNAc on N-glycans**

**MALDI-MS mapping of the N-glycans from acutobin**

The permethylated N-glycans from the purified acutobin afforded a matrix-assisted laser desorption ionization-mass spectrometry (MALDI-MS) profile (Figure 1) similar to that of glycomic mapping at the crude venom level. In the reflectron mode, a major molecular ion signal corresponding to [M + Na]^+ of Manα2Hex1HexNAc1 at m/z 1579 was detected along with several other major signals, which could be assigned as core fucosylated, hybrid-, bi- and triantennary complex-type structures carrying 1–3 terminal NeuAcHex1HexNAc1 epitopes (Figure 1B). Analysis of the same sample in the linear mode showed that the triantennary structure at m/z 4862 was by far the most abundant species, while a minor amount of tetraantennary structure carrying a fourth disialylated antenna could additionally be detected (Figure 1A). Remarkably, heterogeneity in sialylation was represented mostly by species lacking two and not one NeuAc residues.

In support of the terminal disialylated epitope, a prominent signal at m/z 1186 could be assigned as the oxonium ion of NeuAcHex1HexNAc1, arising from MALDI in source prompt fragmentation. Upon treatment with neuraminidase, this fragment ion was no longer afforded by the desialylated sample (Figure 1C). All major molecular ion signals were shifted to m/z values in accordance with the number of NeuAc residues carried, including the complex-type and the pair of hybrid-type structures, as assigned and annotated in Figure 1. Upon further trimming with β-galactosidase and β-HexNAcase, the complex-type structures collapsed into a major signal at m/z 1345, corresponding to a fucosylated trimannosyl core structure (Manα2GlcNAc2Fuc1), whilst the hybrid-type structures yielded core fucosylated Manα2GlcNAc2 and Manα2GlcNAc2 at m/z 1753 and 1549, respectively (data not shown). Alternatively, if desialylation was followed by digestion with α-mannosidase, only Manα,GlcNAc2 and the hybrid-type structures at m/z 1999 and 2203 were affected. 2 and 3 Hex residues were removed from the latter pair, respectively, to give a major signal at m/z 1590, corresponding to (Hexα1HexNAc1)Manα2GlcNAc2Fuc1 (data not shown). Collectively, these data were indicative of common mammalian-type core structures except for the unusually high density of sialylation.

**MS/MS sequencing and identification of disialylated LacNAc terminal epitope**

To determine the distribution of the sialic acids, the major structures were subjected to both low- and high-energy collision-induced dissociation (CID) MS/MS analyses on MALDI-Q/TOF and MALDI-TOF/TOF, respectively, taking advantages of the well-established, complementary fragmentation patterns afforded by the permethylated N-glycans (Yu et al. 2006, 2008). Both the major hybrid- and complex-type N-glycans, as represented by the sodiated precursor ions at m/z 2926 and 4862, respectively, afforded common nonreducing end fragment ions associated with the terminal...
NeuAc$_2$Hex$_1$HexNAc$_1$ epitope (Figure 2). By low collision energy, prominent sodiated B ions were detected at $m$/$z$ 384, 398 and 759, corresponding to (HO)$_1$NeuAc$_1$, NeuAc$_1$ and NeuAc$_2$, which provided a direct evidence for the presence of terminal NeuAc-NeuAc disialyl unit. By high collision energy, this disialyl unit gave a very strong fragment ion signal at $m$/$z$ 717, corresponding to a C$_2$ ion having eliminated a carboxylic acid methyl ester group from the NeuAc. The next C$_3$° and C$_3$ ions at $m$/$z$ 979 and 981, along with E$_3$ and E$_4$ ions at $m$/$z$ 933 and 1137, implicated a direct attachment of the terminal disialyl unit to a Hex that in turn was attached to a HexNAc, thus unambiguously established a terminal NeuAc-NeuAc-Hex-HexNAc- sequence, as shown schematically in Figure 2C (structure III).

Notably, the D ion at $m$/$z$ 1194 informed that the 3-position of the HexNAc was not substituted, whereas the formation of...
Fig. 2. MALDI-MS/MS sequencing of the permethylated N-glycans from acutobin. The hybrid structure (I) carrying one disialyl LacNAc was subjected to both low- and high-energy CID MS/MS on a MALDI-Q/TOF (A) and MALDI-TOF/TOF (B), respectively, whereas the much larger triantennary N-glycan with 3 disialyl LacNAc antenna (II) was analyzed on MALDI-TOF/TOF only (C) since the \( m/z \) of its parent ion exceeded 4000. Full interpretation of the fragment ions was schematically illustrated in structures (I)–(III), in accordance with previously described cleavage patterns and ion nomenclature (Yu et al. 2006, 2008). Structure (III) represents a fully drawn out nonreducing terminal disialyl LacNAc structure, illustrating mostly the nonreducing terminal fragment ions common to both the hybrid and triantennary structures, except the additional cleavage ion produced at the side chain of the internal NeuAc, annotated in red color (\( m/z \) 2503 and 4439 for the hybrid and triantennary structures, respectively). Its formation was similar in nature to the concerted cleavages producing the G ions but specific to internal NeuAc, which was not previously described in the ion nomenclature and not additionally named here.
Site occupancy and glycosylation profile mapping

Although permethylation facilitates MS profiling and unambiguous interpretation of both low- and high-energy CID MS/MS sequencing data of complex glycans, a caveat is that alkali-labile substituents may be lost, while natural O-Me substituent will be rendered cryptic. Analyses of nonderivatized glycopeptides were thus additionally undertaken to uncover any unusual substituents that may be present. It would also allow mapping of the determined glycan structures onto potential glycosylation sites and thus to see whether there is any site-specific glycosylation pattern. Acutobin is known to be translated with amino acid residues 1–18 and 19–24 serving as signal peptide and propeptide, respectively, which would be subsequently cleaved off from the secreted product. Tryptic digest of the mature product is thus predicted to give 19 peptides, named accordingly as T1–T19. The N-glycosylation sites fall within T7, T10 and T19 peptides, with T7 carries two potential sites, making it a total of four potential sites at N101, N105, N124 and N253 (Figure 4).

Direct LC-MS/MS mapping resulted in 66% sequence coverage when database search was performed without specifying enzyme specificity despite the use of sequencing-grade trypsin. This was due mainly to many nontryptic cleavages being induced unexpectedly. Peptides T7 and T10, or their truncated versions still carrying the predicted N-glycosylation sites, could not be identified as nonglycosylated peptides. Neither could T19 but a truncated version of it, ITAGNTTA, was identified as nonglycosylated. If the tryptic digest was first de-N-glycosylated by N-glycosidase F, the sequence coverage rose to 80% due to additional detection of peptides carrying the de-N-glycosylated N101, N105 and N124 sites, as well as T19 carrying de-N-glycosylated N253 site, with corresponding mass shifts of +1 Da due to Asn to Asp conversion. Notably from the isotopic distribution pattern, it was clear that T7 could either be singly or doubly N-glycosylated, which is in consistent with the presence of molecular ion signals corresponding to +1 and +2 Da mass shifts. The results, therefore, indicated that all four potential N-glycosylation sites of acutobin were glycosylated but one of the two closely situated N101/N105 sites and the N253 site near C-terminus did not attain full site occupancy.

The many nonspecific proteolytic cleavages observed rendered interpretation of the glycopeptide LC-MS/MS analysis data difficult. However, based on the calculated m/z values for the expected T7, T10 and T19 and their more abundant nontryptic cleavage products, careful examination of the survey MS profiles and MS/MS spectra led initially to identification of three groups of glycopeptides, corresponding to those sharing a T7, T10 and a truncated T10 peptide cores, the last...
designated as T10a. The main criteria for their grouping together were (1) sharing the same peptide core + HexNAc fragment ion (Y1 ion) as detected in their respective MS/MS spectra and (2) molecular ions afforded were related by glycosyl residue mass differences, with the m/z value of the added glycan moiety corresponded to one of the major glycan compositions defined by glycomic analysis. The latter criterion was extended to include those that were not selected for MS/MS analysis but afforded decent molecular ion signals in the survey MS profiles at the same or neighboring LC elution time points. The range of glycoforms thus delineated for each of these three groups was listed in Table I. Despite the presence of two N-glycosylation sites on T7 that could be both occupied, only the corresponding glycopeptides carrying one and not two N-glycan chains were detected. This was likely due to low abundance and poor ionization efficiency of such large glycopeptides. Similarly, while the truncated T10a carrying biantennary structures could be detected, the only intact T10 glycopeptides detected were those carrying the hybrid-type structures. Finally, the C-terminal T109 glycopeptides were only additionally detected in a separate LC-MS/MS analysis of desialylated sample. However, since nonsialylated complex-type N-glycans were not detected in earlier glycomic mapping, it could be inferred that this N253 site likewise mapping, it could be inferred that this N253 site likewise

Importantly, LC-MS/MS analysis at the nonderivatized glycopeptide level revealed that most of the sialylated glycopeptide signals occurred in pairs of 42 Da apart, corresponding to an acetyl substitution. This pattern was registered for both the T7 and T10 sialylated glycopeptide groups, as shown in Figure 5. It was observed that a majority of the glycopeptides was either not further substituted or carried one degree of acetylation although up to three acetyl increments could be detected for those with triantennary structures (Table I). The location of the extra acetyl group(s) could be readily identified from the characteristic glyco-oxonium ions afforded. Apart from the ubiquitous HexNAc⁺ and HexHexNAc⁺ oxonium ions at m/z 204 and 366, sialylated glycans typically gave fragment ions at m/z 274, 292 and 657, corresponding to NeuAc⁻H₂O, NeuAc⁺ and NeuAcHexHexNAc⁺, respectively. In cases of glycopeptides carrying disialylated LacNAc, additional oxonium ions were expected at m/z 583 and 948 for NeuAc⁺ and NeuAc₂-Hex-HexNAc⁺, respectively. However, oxonium ions were detected instead at m/z 625 and 990, along with ions at m/z 334 and 316 corresponding to an acetylated NeuAc⁺ and that with further loss of a H₂O moiety, for glycopeptides carrying a hybrid-type structure with one additional acetyl group (e.g. the T10a glycopeptide shown in Figure 6A). These ions and the absence of a NeuAcHexHexNAc⁺ ion with an additional acetyl substituent (m/z 699) clearly indicated that the extra acetyl group was carried on the terminal and not the subterminal NeuAc. If the glycopeptides were first treated with mild base prior to similar LC-MS/MS analysis, all the oxonium ions assigned as carrying an extra acetyl group duly disappeared (Figure 6B). In conclusion, the data were consistent with a significant proportion of the terminal disialylated antenna being further substituted by a base labile O-acetyl group on the terminal NeuAc, which was lost upon permethylation and hence not detected by MS mapping of the derivatized glycans despite its apparent abundance revealed by analysis of glycopeptides.

Identification of terminal O-acetylated Neu5Ac(α2-8)Neu5Ac motifs

Fig. 4. Peptide sequence and glycosylation site mapping of acutobin. Eighty percent of the predicted sequence of acutobin (A) could be mapped by LC-MS/MS analysis of its de-N-glycosylated tryptic peptides. Sequences remained not covered were underlined by solid line, which also included the boxed signal peptide and propeptide domains at the N-terminus. Without N-glycosidase F treatment, the sequences underlined with dotted line were additionally not detected, which included the two stretches carrying the predicted N-glycosylation sites. The expected tryptic peptides were named sequentially from the N-terminus, according to the predicted tryptic cleavage sites marked with a vertical line at each of the R and K residues not followed by a proline at its C-terminus. Glycopeptides detected by LC-MS/MS could be mapped to T7, T10a, T10 and T19, as shown in (B), and also highlighted in (A) in black bold prints. The potential N-glycosylation sites at N101, N105, N124 and N253 were highlighted in red.
The high abundance of the characteristic oxonium ions was further taken advantage of in defining the terminal NeuAc-NeuAc linkage by a novel adaptation of the well-established mild periodate oxidation chemistry. It was reasoned that the basis of identifying the intersialic acid linkage by susceptibility of the adjacent OH groups on the side chain of an internal NeuAc involved in a 2–8 linkage could be readily translated to mass shifts and identified by MS/MS analysis. Thus the side chain of an internal NeuAc involved in a 2–8 linkage would be resistant to oxidation but the C8 and C9 of the terminal NeuAc would be lost due to periodate cleavage and served as an internal positive control for the completion of the reaction, as shown in the inset in Figure 6C. In contrast, a 2–9 linkage would preserve two free adjacent OH groups at C7 and C8 of the internal NeuAc and thus be cleaved. Likewise, if an unusual 2–7 linkage exists, then the C8–C9 bond would be cleaved. In each case, different combination of oxonium ions would be produced. In the case of disialylated glycopeptides from acutobin, as exemplified by the same T10a glycopeptide carrying a hybrid-type structure (Figure 6C), the presence of oxonium ions at m/z 292/274 was indicative of an intact NeuAc, whereas those additionally observed at m/z 230/212 corresponded to a C7 analogue of NeuAc having lost its C8–C9 side chain. Notably, instead of m/z 583, 565 and 547, the oxonium ions corresponding to NeuAc were now detected at m/z 521, 503 and 485, respectively. The survival of such a disialyl moiety with an intact internal NeuAc was evident for a resistant 2–8 linkage, whilst the mass shifts were consistent with the side chain of the terminal NeuAc being cleaved off. The same chemistry, whether coupled instead with fluorescent labeling using the 1,2-diamino-4,5-methylenedioxybenzene (DMB) reagent, would be expected to produce an equal amount of a C7 analogue and an intact C9 NeuAc-DMB product, derived respectively from the terminal and internal NeuAc of a NeuAc(α2–8)NeuAc moiety, which could then be detected by fluorometric high performance liquid chromatography (HPLC) (Sato et al. 1998). This was indeed the case when the released glycans were independently subjected to such conventional assay (Figure 6D) and thus confirmed the validity of the alternative MS/MS approach presented here.

**Discussion**

The terminal disialyl motif, Neu5Ac(α2-8)Neu5Ac, identified in this study is thought to have a wide occurrence on mammalian glycoproteins (Sato et al. 2000; Sato 2004; Yasukawa et al. 2005) but due to its low abundance, detailed
characterization of its carrier glycans has been generally lacking. Evidence for its direct attachment to core 1 O-glycans was first obtained on bovine chromogranins (Kiang et al. 1982) and human erythrocyte glycophorins (Fukuda et al. 1987) by MS analysis in the 1980s. More recently, advances in MS-based glycomics have led to occasional identifications of similar disialyl motifs attached to either or both arms of the core 1 O-glycans derived from, for example, mouse natural killer cells (Avril et al. 2006), breast cancer cell lines (Powlesland et al. 2009) and human plasma von Willebrand factor (Canis et al. 2010). In contrast, unambiguous identification of a terminal disialyl motif capping the antenna of multisialylated complex-type N-glycans remains more difficult since it cannot be inferred directly from MS mapping alone. Identification by MS analysis of its presence on usually very large N-glycans relied mostly on positive

Fig. 5. Exemplary MS profiles of detected acutobin glycopeptides encompassing a full range of complex-type N-glycans decorated with terminal disialyl motif with and without additional O-acetylation. Survey MS scans across selected ranges of elution time were summed to show that T10a carried mostly hybrid-type N-glycans with smaller amount of biantennary structure (A), whereas the N-glycans on T7 were mostly multiantennary complex type (B). Additional heterogeneities due to O-acetylation (annotated as +Acn) and the overlapping in elution times of the T7 and T10a glycopeptides could be adequately resolved by careful manual inspection of the individual and summed MS survey scans, as demonstrated by the expanded $m/z$ 1500–1535 mass region shown in (A). The major glycopeptide signals thus detected, assigned and, in most cases, further supported by MS/MS analyses were listed in Table I. The most abundant isotopic peaks are normally the second or third peak in the clusters depending on the $m/z$ values. For clarity and ease of reference to Table I, the $m/z$ values for both the most abundant and the monoisotopic peaks were labeled. A separate LC-MS/MS analysis of the desialylated sample led to additional identification of the T19 glycopeptides, as shown in (C).
detection of the characteristic nonreducing terminal fragment ion, NeuAc2Hex1HexNAc+, which nonetheless did not allow a positive discrimination against another common motif of NeuAc-Hex-(NeuAc)HexNAc, unless supplemented further by linkage analysis. Thus, the reported presence of NeuAc2-8NeuAc2-3Gal-GlcNAc on the N-glycans of human fetal lactosaminoglycans (Fukuda et al. 1984) and human amniotic fluid bronectin (Krusius et al. 1985) by MS analysis remained tentative. In the case of human PA1 embryonal carcinoma lactosaminoglycan where linkages were further verified, it was found that it carried instead NeuAc2-9NeuAc2-3Gal-GlcNAc (Fukuda et al. 1985).

To date, the only fully verified N-glycan structures with terminal Neu5Ac(o2-8)Neu5Ac(o2-3)Gal(o1-4)GlcNAc, apart from those of the acutobin from D. acutus venom reported here, are those from rainbow trout ovarian fluid (Funakoshi et al. 1997), some of which may be additionally fucosylated to give a Neu5Ac(o2-8)Neu5Ac disialylated LeX epitope. However, only the snake venom would ever get into the human body following snake bite and therefore can potentially meddle with the natural human protein–glycan recognition systems. In this context, the glycosylation pattern of acutobin, specifically, and that of D. acutus venom, in general, is truly remarkable in many aspects. Structure-wise, disialylation of each antennae is almost complete, with very little or no monosialylated antennae, in contrast to other reported incidences cited above. Whilst a triantennary complex-type N-glycans appears to be the most abundant structure within the released total glycan pool, hybrid-type structures also constitute a significant glycomic proportion, along with minor amount of bi- and tetraantennary structures. Such a full range of 1–4 disialyl LacNAc antenna carried on discrete structures without the usual heterogeneity associated with sialylation, fucosylation and/or polyLacNAc extension is highly exceptional. With eight sialic acids on a tetraantennary glycan and six sialic acids on a triantennary glycan without polyLacNAc extension, these must rank as N-glycans with highest density of sialylation without carrying polysialic acid chains. More so, all sialic acid...
acids occur in a rather homogeneous setting, presenting a single unique disialyl LacNAc epitope, with O-acetylation serving as the only added variation.

At present, the biological significance of disialylation is poorly understood (Sato 2004). In human and mouse models, the focus of investigation is largely centered on its relationship to developmentally regulated initiation of polysialic acid chains, and its recognition by endogenous glycan-binding proteins. The chief suspects in the latter function are none other than the Siglec family, particularly Siglec 7 expressed strongly on natural killer cells and known to show a preferential binding for Neu5Ac(α2-8)Neu5Ac (Avril et al. 2006; Crocker et al. 2007). However, it should be noted that Neu5Ac(α2-8) Neu5Ac sialoside binding preference is also shared by Siglecs 5, 11, 14 and 16, which collectively represent a wide expression among immune cells, including monocytes, neutrophils, basophils, macrophages, brain microglia, mast cells and dendritic cells (O’Reilly and Paulson 2009). The equally widespread implicated occurrence of the disialyl motif on O-glycans presented by serum (Sato et al. 2001; Yasukawa et al. 2005, 2006, 2007) and brain (Sato et al. 2000, 2002) glycoproteins, in addition to well-known gangliosides carriers of the “b” series, renders conclusive delineation of the endogenous functional implications of Siglec-disialyl motif binding extremely difficult. This is compounded by the fact that the disialyl motif is usually carried only on a small minority of the glycoforms, especially so for the N-glycans, where even firm evidence for occurrence is lacking.

In contrast, acutobin and the crude venom of D. acutus are readily available, since the wild pit vipers are commonly found in south Chinese provinces and also bred in dozens of snake farms. This allowed us not only to fully characterize its N-glycans and glycosylation site occupancy, but, in the process, further develop and optimize highly sensitive MS-based analytical techniques that should be applicable to all future studies associated with defining the disialyl motif. Notably, the high-energy CID MS/MS fragmentation pattern has now been extended to include ions characteristic of permethylated NeuAc-NeuAc-3Hex-4HexNAc. In particular, the highly abundant ion at m/z 717 is extremely useful for high-sensitivity MS/MS validation of the presence of a terminal disialyl motif. We also demonstrated that LC-MS/MS analysis of the nonderivatized glycans at the glycopeptide level can additionally uncover O-acetylation, a common modification of sialic acids, at ease and high sensitivity. The disialyl motif afforded diagnostic oxonium ions that allowed localization of the O-acetyl substituent and, importantly, would facilitate current implementation of targeted glycoproteomics by serving as criteria to efficiently filter out glycopeptide precursors carrying the disialyl modifications. Despite some unwarranted side reactions that may occur at the peptide backbone and complicate assignment, the mass shifts imparted on these highly abundant oxonium ions by periodate oxidation are distinct enough to probe specifically for the disialyl linkage in a one-step process. In each case, the more innovative MS approaches were validated by more conventional chemical analysis. Collectively, we not only established unambiguously the structural details, but also methodologies associated with characterizing them.

The now fully characterized acutobin N-glycans represent an extremely valuable source of naturally occurring, multivalent glycan “standard” carrying the disialyl motif. It will be of interest to see how it may subserve or interfere with a wide range of human immune recognition and activation events mediated by the CD33-related Siglecs. The presence of O-acetylation adds another dimension of complexity as it will either augment or block specific Siglec recognition (Sjoberg et al. 1994; Klein and Roussel 1998), which can nevertheless be manipulated in vitro by alkaline treatment. On the other hand, it is also prudent to evaluate how this unique disialylation may affect the in vivo functioning of acutobin itself upon envenoming, especially in the context of developing it into a better formulated anticoagulant drug for both efficacy and biosafety considerations. In fact, purified and partially purified acutobin, a highly specific fibrinogenase, has been used to treat more than quarter millions of Chinese patients with stroke or other cardiovascular problems, or merely as a stroke-prevention or antithrombotic measure, for the past two decades, without knowing its glycosylation structures. These patients were treated with intravenous injection of acutobin, e.g. within 4–24 h after stroke, to reduce the blood fibrinogen levels to as low as 10% of the normal level (Zhao et al. 2000; Wei et al. 2004), without any apparent short-term adverse side effects or complications.

In general, carbohydrate moieties in venom glycoproteins are postulated to be correlated with the excellent solubility of venom during storage in the venom glands as well as their high stability, bioavailability and fast spreading rate after envenoming. However, the toxinochemical implications of glycan chains of most venom glycoproteins, including acutobin, remain unclear, ambiguous or even controversial. The de-N-glycosylated acutobin retained fibrinogenolytic activities as shown by sodium dodecyl sulfate–polyacrylamide gel electrophoresis analysis of fibrinogen hydrolysis products. We have also produced recombinant wild-type acutobin or its mutants by site-directed mutagenesis from HEK293 cells. Our finding that only N-glycosylation at N253 near the C-terminus of the four sites (Figure 4) was not needed for proper binding of endogenous glycan-binding proteins, including acutobin, derived from envenomation, by reducing their uptake and capping of each of the multiantennary terminus with not only sialylation machinery to ensure an almost complete capping of each of the multiantennary terminus with not only one, but two tandemly linked sialic acids. Although sialylation of otherwise exposed termini does extend the circulatory half life of many serum glycoproteins, including acutobin derived from envenomation, by reducing their uptake and clearance through hepatic asialoglycoprotein receptor (Grewal 2010), this process likewise does not require a disialyl cap per se. Given that such a terminal disialylation is unique for the monotypic D. acutus venom based on our glycomic screening of a wide range of Asian pit viper venom to date, including several species of the Glyoydius, Protobothrops/Trimeresurus, Tropiodolaemus and Ovophis genus (unpublished data), it is intriguing to see whether such a glyco-characteristic confers any advantageous positive selection for adapting to an evolutionary niche. It is anticipated that our ongoing work on the
glycomics of the pit vipers will aid in clarifying the functional implications of evolutionary glycobiology, as well as contributing to better exploitation of a rich sources of snake venoms for biomedical applications.

Materials and methods

Snake venoms and purification of acutobin

Crude venom of D. acutus was obtained from Wu-Long-Mountain snake farm, Ji-Sou, Hunan, China. Acutobin was purified from the venom by ion-exchange chromatography followed by gel filtration on a fast protein liquid chromatography apparatus (Pharmacia, Uppsala, Sweden), exactly as described previously (Wang et al. 2001).

Preparation of glycopeptides and subsequent chemical/enzymatic reactions

Crude venom and purified acutobin samples were solubilized in 50 mM ammonium bicarbonate buffer, pH 8.5, reduced with 55 mM dithiothreitol for 1 h and then alkylated with 65 mM iodoacetamide for 1 h in the dark at 37°C. Excess reagents were removed by passing through a C8 Sep-Pak cartridge (Waters, Milford, MA) before subjecting the recovered protein sample to overnight digestion with sequencing-grade modified trypsin (Promega, Madison, WI) at a 1:20 enzyme:substrate ratio. The resulting tryptic peptides and glycopeptides were desalted by applying to a C18 Sep-Pak cartridge (Waters) pre-equilibrated in 0.1% formic acid and eluted with a step gradient of 25, 50 and 75% acetonitrile in 0.1% formic acid after extensive washing. The peptides/glycopeptides derived from acutobin were then analyzed by nanoLC-MS/MS directly, or after further enzymatic or chemical reactions.

De-O-acetylation was performed by 0.1 N sodium hydroxide treatment at 37°C for 4 h and then desalted by the same C18 Sep-Pak cartridge. Mild periodate oxidation of the de-O-acetylated sample was carried out by adding 15 mM of sodium periodate (Sigma, St Louis, MO) and left at room temperature for 1 h, after which the reaction was terminated by adding 3% ethylene-glycol and similarly desalted by C18 Sep-Pak cartridge. Deimination of the glycopeptides was effected by overnight incubation with 0.2 U neuraminidase from Arthrobacter ureafaciens (Roche, Indianapolis, IN) in 50 mM sodium acetate buffer, pH 6.0, at 37°C. De-N-glycosylation was performed by overnight incubation with N-glycosidase F (Roche) at 37°C, in 50 mM ammonium bicarbonate buffer, pH 8.5. Both enzymatic reactions were followed by similar clean-up using C18 Sep-Pak cartridges prior to nanoLC-MS/MS analysis.

Chemical/enzymatic derivatization of N-glycans and subsequent analyses

N-glycans released by N-glycosidase F were separated from de-N-glycosylated peptides by passing through a C18 Sep-Pak cartridge. Additional exo-glycosidase digestions were carried out under the following conditions: 50 mU α2-3,6,8,9 neuraminidase from A. ureafaciens (recombinant, Escherichia coli, Calbiochem, La Jolla, CA) in 30 μL of 50 mM sodium acetate buffer, pH 5.0; 5 mU β1-4 galactosidase and β1-2,3,4,6 N-acetylglcosaminidase from Streptococcus pneumoniae (recombinant, E. coli, Calbiochem) in 50 μL of 50 mM sodium acetate buffer, pH 6.0; 0.5 U α1-2,3,6 mannosidase from Jack bean (Calbiochem) in 100 μL of 50 mM ammonium acetate buffer containing 1 mM ZnCl₂, pH 5.0. All digestions were carried out at 37°C overnight.

For MALDI-MS analyses, the glycan samples were permethylated using the sodium hydroxide/dimethyl sulfoxide slurry method, as described by Dell et al. (1994). MALDI-MS and MS/MS data were acquired on either a Micromass Q-TOF Ultima™ MALDI (Micromass, Manchester, UK) or a 4700 Proteomics Analyzer (Applied Biosystems, Framingham, MA), exactly as described (Yu et al. 2006). For linkage analysis, the permethylated glycans were further hydrolyzed, reduced and peracetylated to give partially methylated alditol acetates for GC-EI-MS analyses, as described previously (Yu et al. 2008).

Periodate oxidation and C7/C9 Neu5Ac analysis for determination of the disialyl linkage were performed by an adaptation of the method described originally by Sato et al. (1998), preceded by de-O-acetylation in 0.1 N sodium hydroxide at 37°C for 4 h. The alkaline sample solution was adjusted to pH 5.5 with acetic acid before reacted with a final concentration of 15 mM sodium metaperiodate in 100 mM sodium acetate at 4°C for 1 h in the dark, terminated by adding 3% ethylene-glycol. The products were subsequently reduced in 10 mg/mL sodium borohydride/2 M ammonia solution at room temperature for 2 h, desalted by passing through Dowex 50 × 8 (25–50 mesh, H⁺ form) and the borates removed by repeated evaporation in 10% acetic acid/methanol. C7/C9 sialic acid residues were then released by hydrolysis in 0.1 N trifluoroacetic acid at 80°C for 1 h before subjecting to DMB derivatization for reversed-phase C18 HPLC analysis coupled with fluorescence detection, essentially as described previously (Chang et al. 2009).

LC-MS/MS analyses of peptides and glycopeptides

NanoLC-MS/MS analyses of glycopeptides were performed on a Micromass Q-Tof Ultima™ API fitted with a nano-LC sprayer, operated in Precursor Ion Discovery mode, under MassLynx™ 4.0 service packed 2 software control, as described previously (Lin et al. 2008). Briefly, the top five most-intense parent ions detected in the low collision energy (7 eV) MS survey scans were auto-selected for MS/MS analyses only when the corresponding alternating high collision energy (30 eV) MS survey scans registered the presence of characteristic glyco-oxonium ions at m/z 204.084 for HexNac⁺ and m/z 366.139 for HexHexNac⁺. Absence of these two user-specified ions in the triggered MS/MS scans would disqualify the precursor as true positive, and data acquisition would be terminated after the first scan. Otherwise, MS/MS scans would be acquired until the intensity of the precursor ion fell below a predefined threshold (1.5 count/s) or after a maximum of 6 s. Data were manually interpreted for glycopeptide assignment.

For comprehensive sequence coverage of acutobin before and after de-N-glycosylation, tryptic peptides were subjected to nanoLC-MS/MS analyses performed on an LTQ Orbitrap Velos (Thermo Scientific, San Jose, CA) operated in data-dependent mode for automated switching between MS and
MS/MS acquisitions. MS spectra were acquired in the orbitrap in the mass range of m/z 400–2000 with a full-width at half-maximum resolution of 30,000. The top 20 ions were selected for MS² fragmentation by ion trap with an isolation window of 2 Th. The automatic gain control values for MS and MS/MS were set at 500,000 and 10,000, respectively. Raw data files were processed by RAW2MSM 1.1 followed by searching against the Swissprot database (release 2010_09) using the Mascot search engine, with the following considerations for variable modifications: cysteine carbamidomethylation, methionine oxidation, asparagine deamidation. The mass tolerance of precursor was set to 5 ppm and that of fragment ions was 0.6 Da. The enzyme specificity was set as trypsin, or none to consider all nonspecific cleavages.

**Funding**

This work was supported by Academia Sinica (to I.H.T. and K.H.K.).

**Acknowledgements**

Mass spectrometry data acquisition was performed at the NRPGM Core Facilities for Proteomics and Glycomics funded by Taiwan National Science Council.

**Conflict of interest**

None declared.

**Abbreviations**

Ac, acetyl; CID, collision-induced dissociation; DMB, 1,2-diamino-4,5-methylenedioxybenzene; EI, electron impact; ESI, electrospray ionization; GC, gas chromatography; Hex, hexose; HexNAc, N-acetyllactosamine; LaeNAc, N-acetyllactosamine; LC, liquid chromatography; Le², Lewis X; nanoLC, nanoflow LC; NeuAc, N-acetyl neuraminic acid; MALDI, matrix-assisted laser desorption ionization; MS, mass spectrometry; MS/MS, tandem MS; Q/TOF, quadrupole/time-of-flight; Siglecs, sialic acid-binding immunoglobulin-like lectins; TOF/TOF, tandem TOF.

**References**


Nikandrov NN, Deshimaru M, Tani A, Chijiwa T, Shibata H, Chang CC, Fukumaki Y, Ito T, Ohno M. 2005. Purification, primary structures and
evolution of coagulant proteases from Deinagkistrodon actus venom. 


