Mini Review
Published: 29 May, 2018

Journal of Molecular and Cellular Biology Forecast

Glycosylation in Snake Venom Proteinases

Coronado MA*, Eberle RJ and Arni RK*

Multiuser Center for Biomolecular Innovation, Department of Physics, IBILCE/UNESP São Jose do Rio Preto-SP, Brazil

Abstract

One of the most complex post-translational modifications of proteins in eukaryotic cells is glycosylation which not only ensures the correct folding of functional domains but also plays a significant role in biological processes ranging from substrate specificity to ligand recognition.

N-glycosylation is a frequently observed modification in numerous snake venom proteins. The diversity of the *N*-linked carbohydrates present in snake venom proteases can be correlated with toxicity. Snake venom proteases show structural similarities to their mammalian counterparts and hence, are of considerable medical and pharmacological importance. In this review, the general structures of the carbohydrate moieties linked to snake venom serine- and metallo-proteinases are described in detail.

Keywords: Snake veno; Serine proteinase; Metalloproteinase; *N*-,*O*-glycosylation

Abbreviations

ACC: *Agkistrodon contortrix contortrix*; ACLH: pro-*Agkistrodon contortrix laticinctus* hemorrhagic toxin-I; BPA: *Bothrops* protease A; BPTI: Bovine pancreatic trypsin inhibitor; ER: Endoplasmic reticulum; GalNAc: β-D-N-Acetylgalactosamine; GalNAc-Ts: β-D-N-Acetyl-galactosamyltransferases; GlcN: Glucosamine; GlcNAc: β-D-N-Acetylglucosamine; GPI: Glycosyl phosphatidylinositol; Man: β-D-Mannose; Man3-GlcN: Trimannosyl-non-acetylated glucosamine; MP: Metalloproteinase; MS: Mass spectrometry; NMR: Nuclear magnetic resonance; OST: Oligosaccharyltransferase; PCA: Protein C activator; PI: Phosphatidylinositol; PTM: Post translational modification; RER: Rough endoplasmic reticulum; STI: Soybean trypsin inhibitor; SVMP: Snake venom metalloproteinases; SVSP: Snake venom serine proteinases; SVTLE: Snake venom thrombin-like enzymes; TLE: Thrombin-like enzyme

Introduction

Post translational modifications (PTMs) of proteins are ubiquitous in cells and are essential for the manifestation of their diverse activities since they regulate and modulate the functions of proteins. Glycosylation, the post-translational attachment of sugar moieties to proteins provides greater proteomic diversity and extends the sphere of interaction of proteins to encompass a wide range of biological processes including cell attachment, recognition and protein-ligand interactions.

The importance of carbohydrates in biology has moved to the forefront only relatively recently because of several factors that have contributed to the expansion of the field of glycobiology primarily due to the development of novel technologies that enable us to better characterize carbohydrates which has resulted in a significant increase of the number of novel structures identified.

The three main PTMs that involve carbohydrates are N- and O-linked glycosylation and glycosyl phosphatidylinositol (GPI) anchors [1,2]. The commonly encountered monosaccharides and their abbreviations are presented in Figure 1.

Development, pathological events, and cell death are intimately related to glycosylation and protein-linked glycans participate and regulate recognition, protein targeting, cell-cell interactions and binding interactions [1,2].

The oligosaccharide structures between species or even within the same species is extremely diverse and the glycosylation pattern is dependent on the developmental and physiological state of the cell. Studies of various glycoproteins have focused on the thermodynamics of location of sugar moieties on proteins [3,4]. These studies have revealed that the deletion or addition of even a single monosaccharide unit to a protein can significantly influence the equilibrium of the protein between the folded and unfolded states [3,4]. The structural diversity of oligosaccharides encountered in

OPEN ACCESS

*Correspondence:
Monika A. Coronado, Multiuser Center for Biomolecular Innovation, Department of Physics, IBILCE/UNESP São Jose do Rio Preto-SP, Brazil.
E-mail: *monikacoronado@gmail.com*
Raghuvir K. Arni, Monika A. Coronado, Multiuser Center for Biomolecular Innovation, Department of Physics, IBILCE/UNESP São Jose do Rio Preto-SP, Brazil.
Tel: *55-17-3221-2200*
Fax: *+55-17-32212247*
E-mail: *raghuvir.arni@unesp.br*

Received Date: *24 Apr 2018*
Accepted Date: *25 May 2018*
Published Date: *29 May 2018*

Citation: *Coronado MA, Eberle RJ, Arni RK. Glycosylation in Snake Venom Proteinases. J Mol Cell Biol Forecast. 2018; 1(2): 1011.*

ISSN 2643-7953

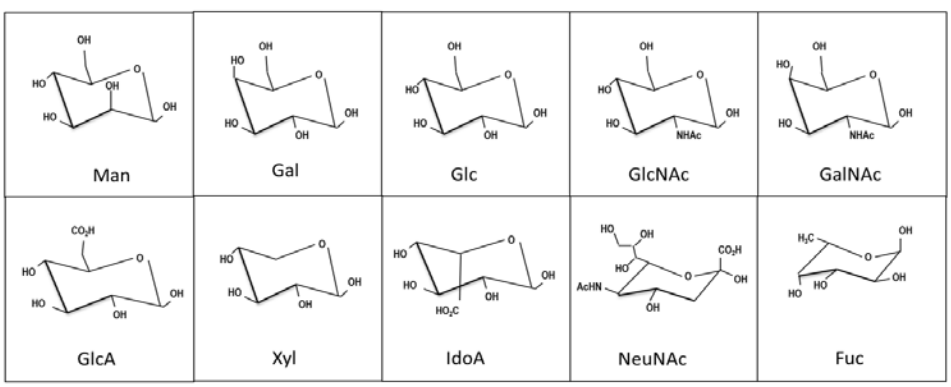

Figure 1: Molecular structures, symbols and abbreviations for monosaccharides. Man (β-D-Mannose), Gal (β-D-Galactose), Glc (β-D-Glucose), GlcNAc (β-D-N-Acetylglucosamine), GalNAc (β-D-N-Acetylgalactosamine), GlcA (β-D-glucuronic acid), Xyl (β-D-Xylose), IdoA (α-L-Iduronic acid), NeuNAc (α-N-Acetylneuraminic acid Sialic Acid), Fuc (α-L-Fucose).

glycoproteins is vast since their biosynthesis is a non-template-driven mechanism involving the synchronized action of many enzymes and also since it is not proofread.

Venom serine proteinases are frequently glycosylated and the carbohydrate moieties are primarily *Asn*-linked [5]. The length of the *N*- or *O*-glycosylation in these proteinases appears to be relevant; however, the correlation between the function and length and variation is in part, still unclear [6,7]. The carbohydrate content may account for 5–30% of the total mass and, in a few cases, the mass of the carbohydrate moiety may exceed that of the protein component [7,8].

The focus of this review is the *N*-glycosylation of snake venom serine- and metalloproteases and their biological aspects.

Overview of Protein Glycosylation

Recognition events are both triggered and modulated by the structure and position of protein-linked glycans [9].This is achieved optimally by oligo- and polysaccharides since they are extremely flexible and can adopt multiple conformational states under physiological conditions [9,10,11]. Unfortunately, this inherent flexibility often limits the use of crystallography for structure determination; hence, we have to resort to the use of other techniques such as mass spectrometry (MS) and Nuclear magnetic resonance (NMR) spectroscopy to determine the structures of carbohydrate moieties [12,13].

Posttranslational modifications are directly dependent on the presence of the relevant enzymes and substrates in the expression cell-line and detailed information of the glycosylation pattern is a prerequisite for understanding the specificities and mechanisms of pharmaceutically relevant glycoproteins, since a connect glycan structure is often essential for their function. Glycoproteins produced in mammalian cell cultures are similar to their human counterparts; however, these are energetically less expensive and functionally less sensitive systems.

Serine proteases that activate protein C in the mammalian coagulation system are generally referred to as Protein C activators (PCAs). The activated form of protein inhibits coagulation via proteolytic inactivation of factors Va and VIIIa. The venom of Southern copper head snake *Agkistrodon contortrix contortrix* (ACC) [15] has potent PCA activity in vitro. Yang and colleagues (2002)

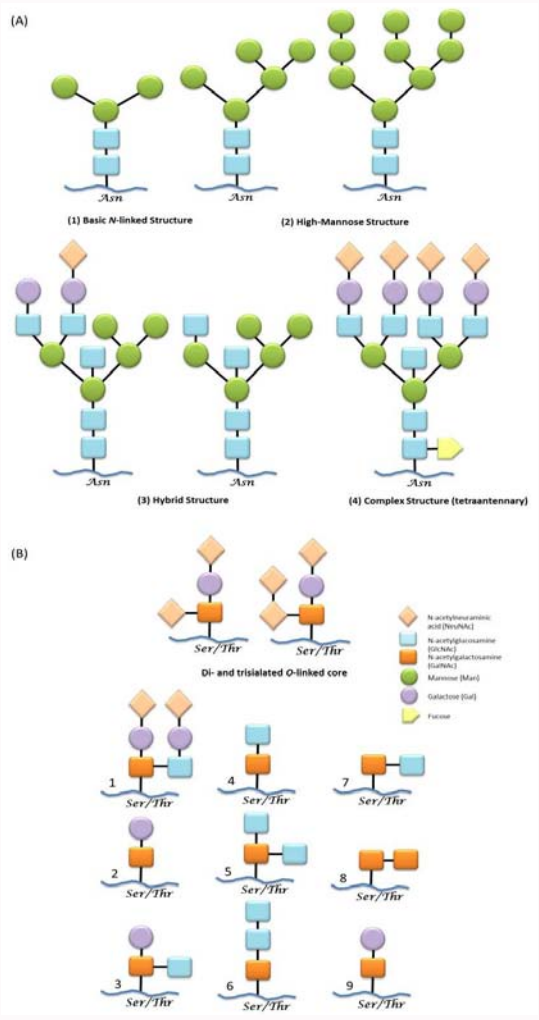

Figure 2: Classification and structure of glycans components. (A) The diverse assortment of *N*-linked glycans is based on the common core pentasaccharide, $Man_3GlcNAc_2$: (1) Basic *N*-linked structure; (2) High-Mannose structure; (3) Hybrid structure; (4) Complex structure. (B) Elongation and termination of *O*-linked glycans is carried out by several glycosultransferase: (1) *O*-linked core 2 Hexasaccharide; (2) Core 1; (3) core 2; (4) core 3; (5) core 4; (6) core 5; (7) core 6; (8) core 7; (9) core 8.

[14] have used yeast *Pichia pastoris* to express the active, glycosylated snake venom serine proteinases (SVSPs): gussurobin, from *Gloydius*

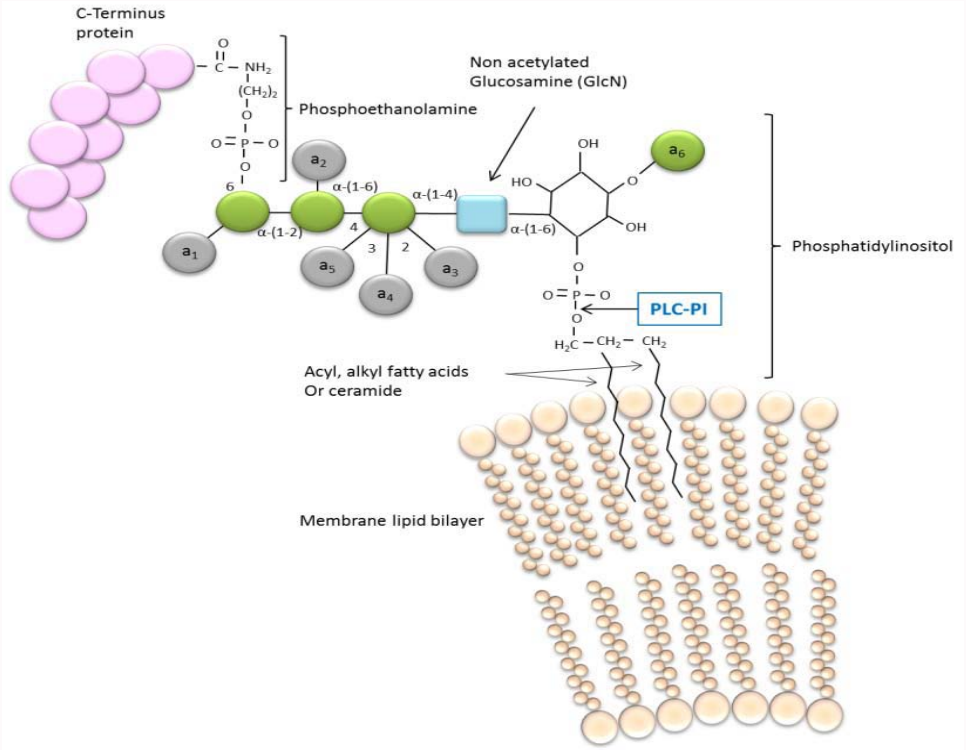

Figure 3: Glycosylphosphatidylinositol (GPI) anchored proteins. Possible modifications to Man_3-GLcN Oligosaccharide core: a_1-Man α-(1-2); a_2-Phosphoethanolamine; a_3- Phosphoethanolamine; a_4- Gal_4; a_5- GalNAcβ-(1-4); a_6- Fatty acid at C_2 or C_3 of inositol.

ussuriensis, and ACC-C from *Agkistrodon contortrix contortrix* [15]. The synthetic gene of PCA based on the ACC protein sequence was inserted into EcoRI–NotI sites of pPICZαA. *Pichia pastoris* strain X33 (Mut+ His+) was used to express recombinant PCA using the AOX1 promoter. *S. Cerevisiae* α-factor signal sequence and *P. pastoris* PHO1 signal sequence were used as signal sequence. The secretion of glycosylated PCA is directed by an α-factor signal sequence. PCA isolated from ACC consists of 231 amino acids and 12 cysteine residues. Recombinant PCA has an apparent molecular weight of ~68 kDa due to a high content of *N*-linked carbohydrates and deglycosylation reduces its size to 28 kDa. This is consistent with sequence prediction that PCA contained three potential *N*-linked glycosylation sites (Asn21, Asn78, and Asn129). Expression in yeast resulted in the addition of a larger carbohydrate moiety to the protein than the naturally occurring snake venom enzyme. [15].

A variety of glycol conjugates are synthesized and the *N*-Linked glycans are linked to the protein backbone via an amide bond to asparagine residues in an Asn-X-Ser/Thr motif, where X represents any amino acid, except Pro. *O*-Linked glycans are linked to the hydroxyl groups of serine or threonine (Figure 2).

GPI-anchored proteins are attached at their carboxy-terminus through a phosphodiester linkage of phosphoethanolamine to a trimannosyl glucosamine core structure (Figure 3) [1,16] via a phosphatidylinositol (PI) group and the reducing end of the latter moiety is bound to the hydrophobic region of the membrane [1,16,17]. Monosaccharide symbols and abbreviations are presented in Figure 1.

In eukaryotic cells, the glycans are synthesized and modified in different compartments equipped with specific enzymes and the initial steps of the biosynthetic pathway of *N*-glycans seem to be similar in practically all organisms and only very minor modifications have been encountered so far [18,19]. The initial steps of the protein *N*-glycosylation occur in the rough endoplasmic reticulum (RER) providing an environment promoting protein folding [18,19]. The RER is the key organelle responsible for the quality control of protein folding and it is on this membrane that the site of the processes production of an elaborate oligosaccharide composed of three glucose, nine mannose and two *N*-acetyl-glucosamine residues is located. In the lumen of the endoplasmic reticulum (ER) lipid-linked oligosaccharides are transferred *en bloc* to specific asparagine residues of the nascent polypeptide (Asn-sequon (N-X-S/T)) by the oligosaccharyltranferase enzyme which is evolutionarily conserved. All these processes are intimately associated with protein folding [20,21] and the biosynthetic processes and the resulting oligosaccharides are evolutionarily conserved in organisms ranging from yeast to mammals.

Classification and Structure of Glycan Components

N-glycosylation

All eukaryotic cells express *N*-linked glycoproteins and protein glycosylation of *N*-linked glycans is actually a co-translational event, occurring during protein synthesis. The characteristic of *N*-glycosylated proteins is a high structural diversity modification present in all domains of life. The acceptor substrate of *N*-glycosylation is an asparagine residue present within the consensus sequence N-X-S/T [1] with a preference for N-X-T over N-X-S sites, and reveals that proline is not tolerated in the second position [1,22]. *Asn-N*-linked glycosylation of proteins by covalent attachment of an oligosaccharide onto the consensus sequence which is generally located in an accessible peptide loop is a fundamental

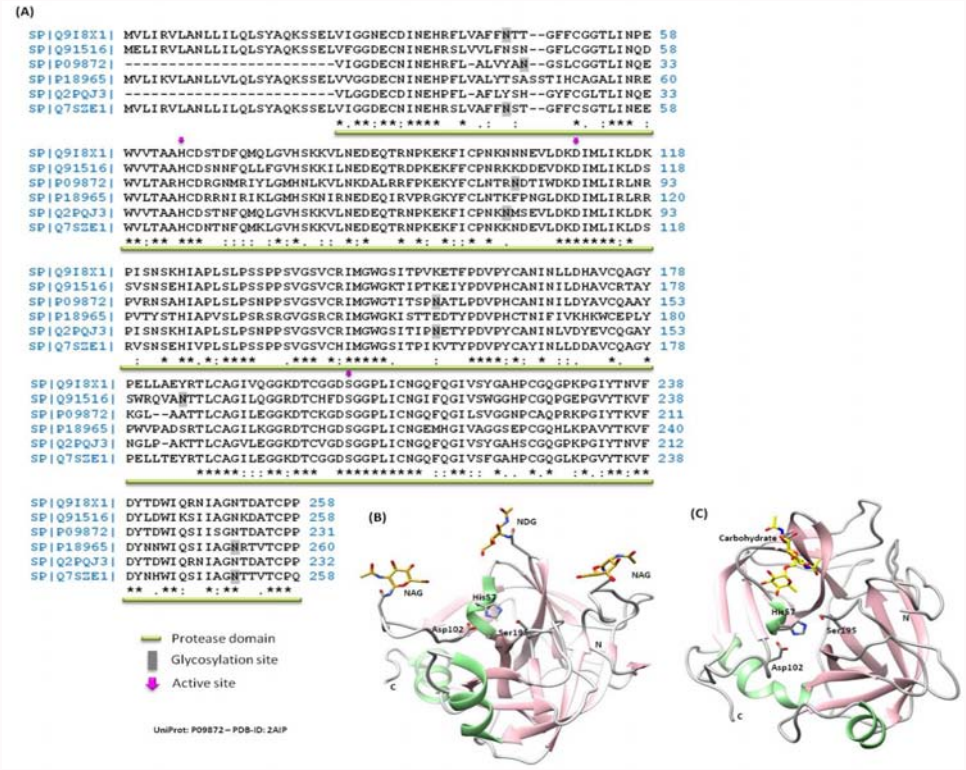

Figure 4: Sequence and structural comparison of SVSP. (A) Multiple sequence alignment of *Deinagkistrodon acutus* (UniProt: Q9I8X1); *Trimeresurus stejnegeri* (UniProt: Q91516); *Agkistrodon contortrix contortrix* (UniProt: P09872); *Daboia siamensis* (UniProt: P18965); *Bothrops jararacussu* (UniProt: Q2PQJ3); *Gloydius saxatilis* (UniProt: Q7SZE1) highlighting the glycosylation sites (grey boxes), (B) Ribbon representations of the crystal structures of ACC-C, the side chains of the amino acids that form the catalytic triad are color-coded (blue, nitrogen; red, oxygen and grey, carbon. The carbohydrate moieties are color-coded (yellow, carbon; red, oxygen; and blue, nitrogen) (NAG - N-Acetyl-D-Glucosamine NDG - 2-(Acetyl amino)-2-Deoxy-A-D-Glucopyranose), (C) Ribbon representations of the model from *Bothrops jararaca*, the side chains of the amino acids that form the catalytic triad are color-coded (blue, nitrogen; red, oxygen and grey, carbon. The carbohydrate moieties are color-coded (yellow, carbon; red, oxygen; and blue, nitrogen).

post-translational modification encountered in eukaryotes and prokaryotes. Intermediate oligosaccharides that are destined for protein incorporation are synthesized by a succession of transferases on the cytoplasmic side of the ER while linked to the dolichol lipid. Subsequently, the addition of a specific number of mannose and glucose molecules occurs. From the dolichol precursor the completed oligosaccharide is then transferred to the Asn of the target glycoprotein by an oligosaccharyltransferase (OST). Additional processing includes trimming of molecules such as glucose and mannose, and the addition of new molecules via transferases in the ER and, to a greater extent, in the Golgi apparatus [1,22]. In the Golgi, high mannose *N*-glycans can be converted in to a variety of hybrid and complex forms (Figure 2A). Inhibition or elimination of glycosylation in the study of *N*-linked glycans can be triggered by a variety of compounds. *N*-glycosylation is highly inhibited in the presence of compactin, coenzyme Q, and exogenous cholesterol. Treatment with tunicamycin completely blocks deglycosylation in that it inhibits GlcNAc C-1-phosphotransferase, which is critical in the formation of the dolichol precursor necessary for synthesizing of *N*-glycans [23-25]. The diverse assortment of *N*-linked glycans is based on the common core pentasaccharide, Man3GlcNAc2. Further processing in the Golgi results in formation of one of the three main classes of *N*-linked glycan sub-types; high-mannose, hybrid, and complex (Figure $2A_{2,3,4}$) and additional monosaccharides may occur in repeating lactosamine units. Additional modifications may include a bisecting GlcNAc at the mannosyl core and/or a fucosyl residue on the innermost GlcNAc. Complex glycans exist in bi-, tri- and tetraantennary shapes (Figure $2A_4$) [24-26]. Stability and solubility are two of the important features of *N*-glycosylation; this modification also protects the protein against proteases. Yeasts often produce very large oligomannosidic glycans but are deficient in the ability to synthesize complex ones.

O-glycosylation

Different types of *O*-linked sugars can be distinguished; O-glycosylation does not begin with the transfer of an oligosaccharide from a dolichol precursor, as in *N*-glycosylation. *O*-linked glycosylation is normally initiated in the Golgi apparatus and is covalently α-linked via its GalNAc moiety to the –OH of serine or a threonine residue of the fully folded and assembled protein within a ternary formed complex, subsequently; a stepwise action of various glycosylation enzymes starting with the UDP-N-Acetyl-Alpha-D-Galactosamine is initiated. A family of 20 different isoforms of polypeptide *N*-Acetyl-galactosamyltransferases (GalNAc-Ts) is involved in the catalysis of the addition of GalNAc onto serine or threonine residues [27]. *O*-Linked glycoproteins are usually large proteins (>200 kDa). *O*-GlcNAc found on nuclear and cytoplasmic proteins are further elongated or modified by acetylation, fucosylation, sialylation, sulfatation, and polylactosamine extension. *O*-Linked glycans tend to be very heterogeneous; hence they are generally classified by their core structure (Figure 2B). Other types of *O*-glycans are: *α*-linked *O*-mannose, *O*-fucose, *β*-linked *O*-xylose, *O*-GlcNAc, *α*- or *β*-linked *O*-galactose, and *α*- or *β*-linked *O*-glucose glycans [28,29].

Anchor glycoproteins

GPI anchored proteins are membrane bound proteins encountered throughout the animal kingdom. GPI anchored proteins are linked at their carboxy terminus through a phosphodiester linkage of phosphoethanolamine to a trimannosyl-non-acetylated glucosamine (Man3-GlcN) core. The reducing end of glucosamine (GlcN) is linked to PI and then anchored via a second phosphodiester linkage to the cell membrane through its hydrophobic region (Figure 3) [16,30]. They contribute to the overall organization of membrane bound proteins and are important in apical protein positioning.

Release of GPI anchored proteins can be accomplished by treatment with Phospholipase C, Phosphatidylinositol specific (PLC-PI) (Figure 3-Blue box). The enzyme specifically hydrolyzes the phosphodiester bond of PI to form a free 1,2-diacylglycerol and glycopeptide-bound inositol cyclic-1,2-phosphate [16].

Isolation purification and analysis

The glycosylation pattern of a specific cell or organism is usually elucidated by the analysis of the carbohydrate structures. As one of the most efficient techniques for protein analysis, SDS-PAGE and 2-dimensional (2D) gel electrophoresis are often used as the first step for the analysis of glycoproteins. Separation of different glycol forms is very intricate due to heterogeneous glycosylation patterns. The glycan structures are released from the protein by specific glycopeptidases followed by labeling with a fluorescent dye. Usually, analysis is then carried out by two-dimensional HPLC and MS methods. Complementary information on the type of glycans attached can be obtained by combining gel electrophoretic analysis and/or lectin probing with treatment by specific exo- and endo-glycosidases as well as respective amidases [31]. Currently, biopharmaceutical glycosylation profiles are most commonly determined on glycans released from the protein backbone either chemically (for example by hydrazinolysis) or enzymatically (*i.e.* using PNGase F). The released glycans are then purified from non-glycan contaminants [32]. The protocol for purification has to be optimized for each kind of tissue separately, owing to the highly complex sample [31,32]. Proteomic analysis by MS such as MALDI-TOF in combination with ESI-MS can provide structural information of the glycans [32-34]. Due to the high complexity, considerable expertise is necessary for MS data analysis.

Snake venom glycoproteins

Snake venoms are complex mixtures of biologically active proteins and the proteinases present in the venoms are classified into trypsin-like serine proteinases and metalloproteinases. These complex mixtures contain an abundance of glycoproteins with *N*-linked carbohydrates, more rarely *O*-linked carbohydrates that can influence activity and specificity of the toxin.

Many of the glycoproteins of snake venoms are enzymes which belong to the serine protease group and affect the hemostatic system via several mechanisms [35-37]. The principal activity of thrombin-like serine proteinases is in fibrin clot formation at the end of the blood clotting cascade. Enzymes able to affect the hemostatic system are divided into four groups: (i) coagulant and procoagulant, (ii) anticoagulant, (iii) inhibitors of platelet function, and (iv), activators of the fibrinolytic system [38]. The literature reports the remarkable biochemical diversity of snake venom serine proteinases. In particular, the reported molecular weights range from 25 to 28 kDa to values as high as 65-70 kDa, although the primary structures of many serine proteinases indicate a molecular weight of < 28 kDa [37,39-42]. This discrepancy suggests a major role of post-translational modification (PTMs) in the molecular diversity of serine proteases, among which glycosylation could be the most prominent [36]. Paes Leme *et al.* (2008), described a trypsin-like serine protease BPA (Bothrops protease A) with a molecular mass of the 25,409 Da but with an apparent molecular weight of 67 kDa obtained by SDS polyacrylamide gel electrophoresis, implying that 62% of the total molecular mass is due to carbohydrate moieties [7].

The snake venom metalloproteinases (SVMPs) are among the most widely studied animal toxins. Several metalloproteinases have been isolated from different snake venoms and their molecular masses range from 20 to 100 kDa. SVMPs are responsible for different systemic effects observed in envenomation which include either pro- or anti-coagulant effects by the proteolysis of plasma components [40]. The metalloproteinases have been grouped into 4 classes: P-I class encompasses the enzymes that only contain a zinc-dependent catalytic domain; P-II class members contain a disintegrin domain in addition to the catalytic domain; P-III class includes the P-II enzyme and the cysteine-rich domain; and P-IV class has an additional lectin-like domain linked by disulfide bonds to a P-III SVMP [40]. SVMPs of all classes cause hemorrhage at the site of envenomation, however, the P-III class enzymes are the more hemorrhagic and contain large carbohydrate moieties [6,40].

Serine proteinases

Snake venom serine proteinases (SVSPs) present *N*- or *O*-glycosylation sites that differ considerably from one another [6]. Potential glycosylation sites may be conserved among snake venom proteins into families. Glycosylation in loops in close proximity to the catalytic sites are most frequently observed and this confers high specificity to the protein which is lost when deglycosylated (Figure 4). Currently, more than 120 amino acids sequences of SVSP are available in the NCBI (http://www.ncbi.nlm.nih.gov/) and UniProt (http://www.uniprot.org/) databanks. Serine peptidase of viperidae snake venom contains the catalytic triad residues, a motif common to serine peptidases of the chymotrypsin family (corresponding to His57, Asp102, and Ser195 in the chymotrypsin based sequence numbering). The sequence alignment of some snake venom serine proteinases and prediction of *N*-glycosylation sites indicates that these are located close to histidine and aspartic acid residues of the catalytic triad (Figure 4A). The figure 4B presents the serine proteinase structure from *Agkistrodon contortrix contortrix* (ACC-C) (PDB ID: 2AIP) and since the carbohydrates are extremely flexible molecules, they may occlude the entrance to active site [43,44].

Recently we have solved the structure of a serine proteinase from *Bothrops jararaca* HS114 (to be publish). The model presents just one glycosylated site (Asn35) composed with N-acetyl-D-Glucosamine - NAG and Alpha-L-Fucose –FUC as presented in figure 4C.

ACC-C contains 16% carbohydrate *N*-linked, the consensus signal sequence for the attachment of carbohydrate moieties was identified at positions 38, 96a, and 148 [44]. Clear electron density map of a thrombin-like protein from *Agkistrodon contortrix contortrix* venom indicated *N*-glycosylation at positions Asn96a and Asn148 (Figure 4B) [45-47]. Some glycosylation sites are present close to the active site and probably play a role in restricting access of large ligand molecules such STI and BPTI [43-45]. Thus, it appears that glycosylation in serine proteases creates steric hindrance and plays a role in selectivity [43-45]. *Agkistrodon acutus* possess

Figure 5: Sequence and structural comparison of SVMP. (A) Sequence alignment of metalloproteinase P-III Class with the respective glycosylation site shown in grey boxes. *Crotalus atrox* (Uniprot: Q9DGB9 and Q90282); *Daboia siamensis* (Uniprot: Q7LZ61). (B) Ribbon diagrams of VAP2, the carbohydrate moieties are color-coded (yellow, carbon; red, oxygen; and blue, nitrogen). The carbohydrate moiety linked to N371 is shown as a stick representation (BMA: Beta-D-Mannose, NAG: N-Acetyl-D-Glucosamine, FUC: Alpha-L-Fucose). The M-domain, disintegrin and cysteine-rich domain are shown in green, pink, and grey, respectively. Calcium ions are represented as orange spheres.

two serine proteinase isoform, AaV-SP-I and AaV-SP-II both of which contain an *N*-linked carbohydrate group at position 35 and these *N*-linked carbohydrate interfere spatially with the binding of some macromolecular inhibitors, and thus, may be involved in the regulation of enzyme-inhibitor interactions and catalytic activities [44-46].

Interestingly, glycosylation lends greater thermal stability to the thrombin-like enzyme (TLE) isolated from *B. jararacussu* venom (BJ-48). However, the deglycosylation increases susceptibility to trypsin inhibitors such as soybean trypsin inhibitor (STI) [36]. In the case of *Bothrops* protease A (BPA), isolated from *B. jararaca* venom, the presence of glycosylation is responsible for the observed thermal stability. Serine proteinase inhibitors such as STI, bovine pancreatic trypsin inhibitor (BPTI), and antithrombin III fail to inhibit BPA, probably due to the steric hindrance caused by the carbohydrate moieties, however, inhibition by small inhibitors such as benzamidine [7,39,47] is not affected. The susceptibility of the enzyme to protein inhibitors is not increased by partial deglycosylation of BPA but activity toward both fibrinogen and D-Val-Leu-Arg-pNA was higher [7]. Deglycosylation indicates a functional trade-off between optimal enzyme activity and in vivo stability, showing that the protective effect of glycosylation is not without a price in terms of enzyme efficiency.

SVSPs can be easily purified from crude venoms by a variety of chromatographic techniques. Generally, the process of separation starts with a fractionation based on size exclusion chromatography followed by either ion-exchange and/or affinity binding (i.e. benzamidine sepharose) chromatography [48,49].

Metalloproteinases

Degradation of the components of the basement membranes underlying capillary endothelial cells is the essentially action of the SVMP and is caused by the disruption of the cell vessel wall, allowing the leakage of the contents into the stroma. The hemorrhagic activity is among the major lethal factors in viper snake venom bites. However, certain SVMPs do not possess hemorrhagic activity, but trigger the homeostasis mediated by pro- or anti-coagulant effects. The molecular weights of metalloproteinases from snake venoms can vary from 20 to 100 kDa and belong to different classes according to their domain organization [40,50].

BmooMPα-I from *Bothrops moojeni*, belongs to the P-I class of SVMPs and possesses fibrinogenolytic activity [51,52]. BaP1 from *Bothrops asper* presents hemorrhagic, inflammatory and myonecrotic activities [53,54]. On the other hand, VaF1, a P-III fibrinogenolytic metalloproteinase from *Vipera ammodytes ammodytes* venom, presents two consensus *N*-glycosylation sites that are present in the sequence of two sites, in the MP domain and/or in the cysteine-rich domain. The extent of its glycosylation is low, representing only 5.2% of the total molecular mass of the protein [55]. Oliveira *et al.* (2010), described the structural effect of glycosylation in four SVMPs from *Bothrops jararaca* [56]. One of the proteins is HF3 a P-III class metalloproteinase which is highly glycosylated is approximately 80 times more haemorrhagic than bothropasin (P-III class) which has a minor carbohydrate moiety. The degradation profiles of fibrinogen, fibronectin, vitronectin, von Willebrand factor, collagens IV and VI, laminin, and Matrigel™; present a different profile for diverse metalloproteinase P-I and P-III classes, however, bothropasin (P-III) was also able to degrade collagen I. Two of these metalloproteinases, bothropasin and BJ-PI lose their structural stability when they are *N*-deglycosylated. Nevertheless, HF3 remained intact, although it's haemorrhagic and fibrinogenolytic activities were partially affected [56]. The carbohydrate moieties of highly glycosylated, hemorrhagic SVMP may play a role in their stability and interaction with key substrates of the extracellular matrix *in vivo*.

HF3 is a highly glycosylated P-III class proteinase composed of 416 residues, with a calculated molecular mass of approximately 46 kDa. Including the five *N*-glycosylation sites, the molecular mass was estimated by SDS-PAGE to be 70 kDa, implying that about 35% of its molecular mass assessed by SDS-PAGE is due to carbohydrate moieties. Bothropasin is far less haemorrhagic, although highly proteolytic, P-III class proteinase that contains 421 residues (approx. molecular mass 46.6 kDa), including one *N*-glycosylation site, the estimated SDS-PAGE shows approx. 48 kDa [57]. Catrocollastatin/vascular apoptosis-inducing protein VAP2 is a metalloproteinase from *Crotalus atrox*, possessing Metalloprotease/Disintegrin-like/Cysteine-rich (MDC) domains is a single chain protein with a molecular mass of 55 kDa and is composed of 419 amino acids. The protein bears the classic domain architecture of a disintegrin and metalloproteinase ADAM (A Disintegrin and Metalloproteinase) family protein (Figure 5A and B).

Native pro-*Agkistrodon contortrix laticinctus* hemorrhagic toxin-I (ACLH) was tested with five lectins for specific binding to carbohydrate moieties and indicates sialic acid terminally α(2–6)-linked to galactose. Deglycosylation of native ACLH produced a shift in the mobility in SDS-PAGE (approximately 10% in molecular weight), indicating that post-translational glycosylation contribution to the total protein molecular mass and native and deglycosylated enzymes exhibited similar kinetic parameters [58]. Garcia and colleagues (2004), showed that ACLH has sialic acid terminally α(2–6)-linked to galactose in a complex *N*-glycan chain and the recombinant, non-glycosylated and native proteins hydrolyzed the fluorophore substrate with evident differences in their kinetic parameters [58]. The hydrodynamic volume occupied by the carbohydrate moieties is significantly larger as many studies suggest and is probably important for the observed protein conformation and folding [43,58].

Structural studies

Since glycosylation is chemically and structurally heterogeneous, it can interfere with the formation of single crystals suitable for structural studies. Glycoproteins can be crystallized in a wide range of conditions and trials are usually set up using similar crystallization screens as non-glycosylated proteins. Many reviews have addressed this problems related to glycoprotein crystallization such as the increase in surface entropy associated with large post-translational modifications [59]. However, it is to be noted that the presence of glycans can be an advantage for crystallization as they can form essential intermolecular contacts in crystal lattices [60] and crystallization trials with more than one glycoform can be advantageous.

Due to the flexibility of the glycan chains, the X-ray crystal structures are only able to determine the positions and structures of only the initial GlcNAc residue even if additional sugar residues are attached to the protein since the electron density for additional, flexible, sugar residues is often diffuse or is even totally absent. In rare cases, more of the oligosaccharide chain can be observed in electron density maps when it is involved in forming either crystal contacts or when it is interacting with the polypeptide chain, thus, restricting the flexibility of the carbohydrate chain [61].

Function

Post-translation modifications play a key role in structure-function properties of the proteins or enzymes. Pos-translational attachment of carbohydrates to proteins is known as glycosylation which takes place primarily in the endomembrane system. Most of the snake venom thrombin-like enzymes are attached through N-glycosylation. It has been well documented that glycosylation is required for the proper folding and efficient secretion of many, but no all glycoprotein. Hence, the importance of the carbohydrate on protein folding, secretion or for any other purpose must be determined for each individual glycosylated protein. The exact role of glycosylation on snake venom thrombin-like enzymes (SVTLEs) is still unclear and many incomplete and conflicting data are available. Some of the predicted functions are: thermostability that protects the enzyme from thermal denaturation [36]; catalytic resistance against extreme pH and temperature [62]; role of glycosylation against inhibition of enzyme activity. Some *N*-linked carbohydrate residue(s) may be involved in steric blockage of a particular inhibitor to the enzyme active site [36]; role of glycosylation on alteration of catalytic activity of the enzyme. Many studies have demonstrated that glycosylation improves the catalytic efficiency of SVTLEs towards the macromolecular substrates such as fibrinogen [36,39,42,47]. However, PaesLeme *et al.* 2008, observed that partial removal of *N*-linked carbohydrate residues from BPA, a SVTLE from *Bothrops jararaca* resulted in increase in its fibrinogenolytic activity compared to native, the partial deglycosylation may have facilitated its interaction with fibrinogen [7].

Conclusions

Many critical biological processes such as cell adhesion, adaptive, innate immunity and cell signaling depend on the ability of proteins to identify and differentiate between carbohydrate molecules. The structure determination of flexible and dynamic molecules is a great challenge and 3D structural characterization of carbohydrate–protein complexes is required to obtain a better understanding of the structural and dynamic properties of carbohydrates. Crystallography and NMR spectroscopy have added the first rungs of the technical ladder that leads to high throughput methods for carbohydrate structure determination. Experimental and computational methods offer a new, alternative path to determining the 3D structures of carbohydrate–protein complexes by characterizing the interaction of interfaces between sugar and protein molecules. NMR strategies that use partially oriented samples, in combination with computational simulations, are a promising path for 3D carbohydrate–protein structure determination. Considerable expansion in glycomics capabilities is required to fully exploit protein glycosylation for the development of the next generation of therapies using those proteins.

The analysis of snake venom glycosylation is an emerging field in glycobiology and a number of structures have been determined recently and several completely new glycan modifications have been characterized.The investigation of glycosylation capacities combined with the analysis of substrate specificity is a growing field in this area and promises new insights.

Funding

This research was supported by grants from the CNPq (Science without Frontiers), FAPESP and CAPES.

References

1. Marshall RD. Glycoproteins. Annu. Rev. Biochem. 1972; 41: 673-702.
2. Rademacher TW, Parekh RB and Dwek RA. Glycobiology. Annu. Rev. Biochem. 1988; 57: 785-838.
3. Shental-Bechor D and Levy Y. Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. PNAS. 2008; 105: 8256-8261.
4. Shental-Bechor D and Levy Y. Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. Curr. Opin. Struct. Biol. 2009; 19: 524-533.
5. Tanaka N, Nakada H, Itoh N, Mizuno Y, Takanishi M, Kawasaki T, et al. Novel structure of the N-acetylgalactosamine containing N-glycosidic carbohydrate chain of batroxobin, a thrombin-like snake venom enzyme. J. Biochem. 1992; 112: 68-74.
6. Serrano SM and Maroun RC. Snake venom serine proteinases: sequence homology vs. substrate specificity, a paradox to be solved. Toxicon. 2005; 45: 1115-1132.
7. PaesLeme AF, Prezoto BC, Yamashiro ET, Bertholim L, Tashima AK, Klitzke CF, et al. Bothrops protease A, a unique highly glycosylated serine proteinase, is a potent, specific fibrinogenolytic agent. J Thromb Haemost. 2008; 6: 1363-1372.
8. Jin Y, Lu QM, Wei JF, Li DS, Wang WY and Xiong YL. Purification and characterization of jerdofibrase, a serine protease from the venom of Trimeresurus jerdonii snake. Toxicon. 2001; 39: 1203-1210.
9. Kelleher DJ and Gilmore R. An evolving view of the eukaryotic oligosaccharyltransferase. Glycobiology. 2006; 16: 47R-62R.
10. Varki A, Esko JD and Colley KJ. Cellular Organization of Glycosylation. In Essentials of Glycobiology, 2nd edn (Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW and Etzler ME, eds), Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA. 2009; pp. 37-46.
11. DeMarco ML and Woods RJ. Structural glycobiology: a game of snakes and ladders. Glycobiology. 2008; 18: 426-440.
12. Mayer M and Meyer B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. Angew. Chem. Int. Ed. Engl. 1999; 38: 1784-1788.
13. Mayer M and Meyer B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. J. Am. Chem. Soc. 2001; 123: 6108-6117.
14. Yang Q, Hu XJ, Xu XM, An LJ, Yuan XD, Su ZG, et al. Cloning, expression and purification of gussurobin, a thrombin-like enzyme from the snake venom of Gloydius ussuriensis. ABBS. 2002; 34: 6-10.
15. Kunes YZ, Sanz MC, Tumanova I, Birr CA, Shi PQ, Bruguera P, et al. Expression and characterization of a synthetic protein C activator in Pichia pastoris. Protein Express Purif. 2002; 26: 406-415.
16. Ferguson MAJ, Kinoshita T and Hart GW. Glycosylphosphatidylinositol Anchors. In Essentials of Glycobiology, 2nd edn (Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W. and Etzler, M.E., eds), Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA. 2009; pp. 143-162.
17. Campbell AS. Glycosylphosphatidylinositols. In Glycoscience Chemistry and Chemical Biology I–III, 1st edn, Volume 2, (Fraser-Reid, B., Tatsuta, K. and Thiem, J., eds), Springer Science+Business Media S.A: Berlin, Germany. 2001; pp. 1695-1777.
18. Kornfeld R and Kornfeld S. Assembly of asparagine-linked oligosaccharides. Annu. Rev. Biochem. 1985; 54: 631-664.
19. Aebi M. *N*-linked protein glycosylation in the ER. Biochim. Biophys. Acta. 2013; 1833: 2430–2437.
20. Roth J. Protein *N*-glycosylation along the secretory pathway: relationship to organelle topography and function, protein quality control, and cell interactions. Chem. Rev. 2002; 102: 285-304.
21. Roth J, Zuber C, Park S, Jang I, Lee Y, Kysela KG, et al. Protein N-glycosylation, protein folding, and protein quality control. Mol. Cells. 2010; 30: 497-506.
22. Shrimal S, Cherepanova NA and Gilmore R. Cotranslational and posttranslocational N-glycosylation of proteins in the endoplasmic reticulum. Semin. Cell Dev. Biol. Academic Press. 2010; 41: 71-78.
23. Merlie JP, Sebbane R, Tzartos S and Lindstrom J. Inhibition of glycosylation with tunicamycin blocks assembly of newly synthesized acetylcholine receptor subunits in muscle cells. J. Biol. Chem. 1982; 257: 2694-2701.
24. Stanley P, Schachter H and Taniguchi N. N-Glycans. In Essentials of Glycobiology, 2nd edn (Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W. and Etzler, M.E., eds), Cold Spring Harbor Laboratory Press: Cold Spring Harbor NY, USA. 2009; pp. 101-114.
25. Stanley P and Commings RD. Structures common to Different Glycans. In Essentials of Glycobiology, 2nd edn (Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W. and Etzler, M.E., eds), Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA. 2009; pp. 175-198.
26. Sagi D, Peter-Katalinic J, Conradt HS and Nimtz M. Sequencing of tri- and tetraantennary N-glycans containing sialic acid by negative mode ESI QTOF tandem MS. J. Am. Soc. Mass. Spectrom. 2002; 13: 1138-1148.
27. Bennett EP, Mandel U, Clausen H, Gerken TA, Fritz TA and Tabak LA. Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. Glycobiology. 2012; 22: 736-756.
28. den Steen PV, Rudd PM, Dwek RA and Opdenakker G. Concepts and principles of O-linked glycosylation. Crit. Rev. Biochem. Mol. Biol. 1998;

33: 151-208.

29. Tran DT and Ten Hagen KG. Mucin-type O-glycosylation during development. J. Biol. Chem. 2013; 288: 6921-6929.

30. Rudd PM, Mattu TS, Zitzmann N, Mehta A, Colominas C, Hart E, et al. Glycoproteins: rapid sequencing technology for N-linked and GPI anchor glycans. Biotechnol Genet Eng Rev. 1999; 16: 1-22.

31. Geyer H and Geyer R. Strategies for analysis of glycoprotein glycosylation. Biochim. Biophys. Acta-Proteins and Proteomics. 2006; 1764: 1853-1869.

32. Domann PJ, Pardos-Pardos AC, Fernandes DL, Spencer DIR, Radcliffe CM., Royle L, et al. Separation-based Glycoprofiling Approaches using Fluorescent Labels. Proteomics. 2007; 7: 70-76.

33. Hamouda H, Kaup M, Ullah M, Berger M, Sandig V, Tauber R, et al. Rapid analysis of cell surface N-glycosylation from living cells using mass spectrometry. J. Proteome Res. 2014; 13: 6144-6151.

34. Chen W, Smeekens JM and Wu R. Comprehensive Analysis of Protein N-Glycosylation Sites by Combining Chemical Deglycosylation with LC–MS. J. Proteome Res. 2014; 13: 1466-1473.

35. Mackessy SP. Thrombin-Like Enzymes in Snake Venoms. In Toxins and Hemostasis, 1st edn (Kini, R.M., Clemetson, K.J., Markland, F.S., McLane, M.A., Morita, T., eds), Springer Science+Business Media B.V. Dordrecht, Netherlands. 2010; pp. 519-557.

36. Silva-Junior FP, Guedes HL, Garvey LC, Aguiar AS, Bourguignon SC, Di Cera E, et al. BJ-48, a novel thrombin-like enzyme from the Bothrops jararacussu venom with high selectivity for Arg over Lys in P1: role of N-glycosylation in thermostability and active site accessibility. Toxicon. 2007; 50: 18-31.

37. Pradniwat P and Rojnuckarin P. Snake venom thrombin-like enzymes. Toxin Rev. 2013; 33: 16-22.

38. Markland FS. Snake venoms and the hemostatic system. Toxicon. 1998; 36: 1749-1800.

39. Murayama N, Saguchi KI, Mentele R, Assakura MT, Ohi H, Fujita Y, et al. The unusual high molecular mass of Bothrops protease A, a trypsin-like serine peptidase from the venom of Bothrops jararaca, is due to its high carbohydrate content. Biochim. Biophys. Acta-Proteins and Proteomics. 2003; 1652: 1-6.

40. Fox JW and Serrano SM. Insights into and speculations about snake venom metalloproteinase (SVMP) synthesis, folding and disulfide bond formation and their contribution to venom complexity. FEBS J. 2008; 275: 3016-3030.

41. Menaldo DL, Bernardes CP, Santos-Filho NA, de Andrade Moura L, Fuly AL, Arantes EC, et al. Biochemical characterization and comparative analysis of two distinct serine proteases from Bothrops pirajai snake venom. Biochimie. 2012; 94: 2545-2558.

42. Mukherjee AK. The pro-coagulant fibrinogenolytic serine protease isoenzymes purified from Daboia russelii russelii venom coagulate the blood through Factor V activation: role of glycosylation on enzymatic activity. PloS one. 2014; 9: e86823.

43. Murakami MT and Arni RK. Thrombomodulin-independent Activation of Protein C and Specificity of Hemostatically Active Snake Venom Serine Proteinases: crystal structures of native and inhibited Agkistrodon contortrix contortrix protein C activator. J. Biol. Chem. 2005; 280: 39309-39315.

44. Giuseppe PO, Souza TACB, Murakami MT, Betzel C, Arni RK. Protein C Activators from Snake Venom. In Handbook of Proteolytic Enzymes, 3rd edn, Volume 1 (Rawlings, N.D. and Salvesen, G.S., eds), Elsevier: San Diego, USA. 2012; pp. 3046-3047.

45. Zhu Z, Liang Z, Zhang T, Zhu Z, Xu W, Teng M, et al. Crystal structures and amidolytic activities of two glycosylated snake venom serine proteinases. J. Biol. Chem. 2005; 280: 10524-10529.

46. Kang TS, Georgieva D, Genov N, Murakami MT, Sinha M, Kumar RP, et al. Enzymatic toxins from snake venom: structural characterization and mechanism of catalysis. FEBS J. 2011; 278: 4544–4576.

47. Botos I and Wlodawer A. The expanding diversity of serine hydrolases. Curr. Opin. Struct. Biol. 2007; 17: 683-690.

48. Ullah A, Souza TACB, Zanphorlin LM, Mariutti RB, Santana VS, Murakami MT, et al. Crystal structure of Jararacussin-I: The highly negatively charged catalytic interface contributes to macromolecular selectivity in snake venom thrombin-like enzymes. Prot. Struc. Report. 2012; 22: 128-132.

49. de Oliveira LMF, Ullah A, Masood R, Zelanis A, Spencer PJ, Serrano SMT, et al. Rapid purification of serine proteinases from Bothrops alternatus and Bothrops moojeni venoms. Toxicon. 2013; 76: 282-290.

50. Fox JW and Serrano SM. Structural considerations of the snake venom metalloproteinases, key members of the M12 reprolysin family of metalloproteinases. Toxicon. 2005; 45: 969-985.

51. Bernardes CP, Santos-Filho NA, Costa TR, Gomes MS, Torres FS, Costa J, et al. Isolation and structural characterization of a new fibrin(ogen)olytic metalloproteinase from Bothrops moojeni snake venom. Toxicon. 2008; 51: 574–584.

52. Souza TACB, Giuseppe PO, Arni RK and Murakami MT. BmooMPα-I (Bothrops moojeni). In Handbook of Proteolytic Enzymes, 3rd edn, Volume 1 (Rawlings, N.D. and Salvesen, G.S., eds), Elsevier, San Diego, USA. 2013: 1000-1005.

53. Watanabe L, Shannon JD, Valente RH, Rucavado A, Alape-Giron A, Kamiguti AS, et al. Amino acid sequence and crystal structure of BaP1, a metalloproteinase from Bothrops asper snake venom that exerts multiple tissue-damaging activities. Protein Sci. 2003; 12: 2273-2281.

54. Gutiérrez J, Romero M, Díaz C, Borkow G and Ovadia M. Isolation and characterization of a metalloproteinase with weak hemorrhagic activity from the venom of the snake Bothrops asper (terciopelo). Toxicon. 1995; 33: 19-29.

55. Leonardi A, Sajevic T, Latinović Z, Pungerčar J, Balija ML, Bakija AT, et al. Structural and biochemical characterisation of VaF1, a P-IIIa fibrinogenolytic metalloproteinase from Vipera ammodytes ammodytes venom. Biochimie. 2015; 109: 78-87.

56. Oliveira AK, Leme AP, Asega AF, Camargo ACM, Fox JW and Serrano SM. New insights into the structural elements involved in the skin haemorrhage induced by snake venom metalloproteinases. Thromb. Haemost. 2010; 104: 485-497.

57. Assakura MT, Silva CA, Mentele R, Camargo AC and Serrano SM. Molecular cloning and expression of structural domains of bothropasin, a P-III metalloproteinase from the venom of Bothrops jararaca. Toxicon. 2003; 41: 217-227.

58. García LT, Parreiras e Silva LT, Ramos OHP, Carmona AK, Bersanetti PA and Selistre-de-Araujo HS. The effect of post-translational modifications on the hemorrhagic activity of snake venom metalloproteinases. Comp. Biochem. Physiol. 2004; 138: 23–32.

59. Chang VT, Crispin M, Aricescu AR, Harvey DJ, Nettleship JE, Fennelly JA, et al. Glycoprotein structural genomics: solving the glycosylation problem. Structure. 2007; 15: 267-273.

60. Mesters JR and Hilgenfeld R. Protein Glycosylation, Sweet to Crystal Growth? Cryst. Growth Des. 2007; 7: 2251-2253.

61. Crispin M, Scanlan CN and Bowden TA. Structure and Biosynthesis of Glycoprotein Carbohydrates. In Comprehensive Biotechnology, 2nd edn, Volume 1- Moo-Young, M., edElsevier. Amsterdam, Netherlands. 2011: 73-90.

62. Costa FLS, Rodrigues RS, Izidoro LFM, Menaldo DL, Hamagichi A, Homsi-Brandeburgo MI, et al. Biochemical and functional properties of a thrombin-like enzyme isolated from Bothrops pauloensis snake venom. Toxicon. 2009; 54: 725-735.