

# Sequence and phylogenetic analysis of viper venom serine proteases

Sakthivel Vaiyapuri<sup>1\*</sup>, Nethaji Thiyagarajan<sup>2</sup>, E Gail Hutchinson<sup>1</sup> & Jonathan M Gibbins<sup>1</sup>

<sup>1</sup>Institute for Cardiovascular and Metabolic Research, University of Reading, Reading, United Kingdom; <sup>2</sup>Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom; Sakthivel Vaiyapuri - Email: s.vaiyapuri@reading.ac.uk; Phone: +441183787047; Fax: +441183787045; \*Corresponding author

Received July 19, 2012; Accepted August 02, 2012; Published August 24, 2012

## Abstract:

Snakebites are a major neglected tropical disease responsible for as many as 95000 deaths every year worldwide. Viper venom serine proteases disrupt haemostasis of prey and victims by affecting various stages of the blood coagulation system. A better understanding of their sequence, structure, function and phylogenetic relationships will improve the knowledge on the pathological conditions and aid in the development of novel therapeutics for treating snakebites. A large dataset for all available viper venom serine proteases was developed and analysed to study various features of these enzymes. Despite the large number of venom serine protease sequences available, only a small proportion of these have been functionally characterised. Although, they share some of the common features such as a C-terminal extension, GWG motif and disulphide linkages, they vary widely between each other in features such as isoelectric points, potential N-glycosylation sites and functional characteristics. Some of the serine proteases contain substitutions for one or more of the critical residues in catalytic triad or primary specificity pockets. Phylogenetic analysis clustered all the sequences in three major groups. The sequences with substitutions in catalytic triad or specificity pocket clustered together in separate groups. Our study provides the most complete information on viper venom serine proteases to date and improves the current knowledge on the sequence, structure, function and phylogenetic relationships of these enzymes. This collective analysis of venom serine proteases will help in understanding the complexity of envenomation and potential therapeutic avenues.

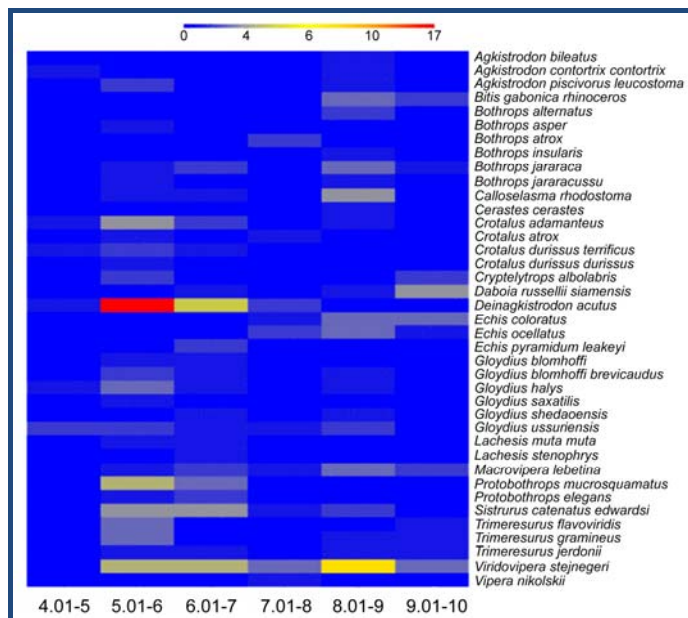
## Background:

Snakebite is a major neglected public health issue particularly among agricultural communities living in rural regions throughout the world [1, 2]. An estimated 2.5 million people are bitten by snakes each year and these are estimated to result in up to as high as 95000 deaths worldwide [2, 3]. Snake venoms are complex mixtures of enzymatic [4, 5] and non enzymatic proteins [6], together with other components such as carbohydrates, lipids, nucleosides and metals [4, 7]. Snake venom serine proteases are major components and have been identified mainly in the venoms of snakes belonging to the viperidae family with a few occurring in members of the elapidae, colubridae and hydrophidae families [8]. Many snake venom serine proteases exert their effects through the ability to disrupt the normal haemostasis of envenomed prey and victims [9]. Indeed, viper venom serine proteases (VVSPs) affect various

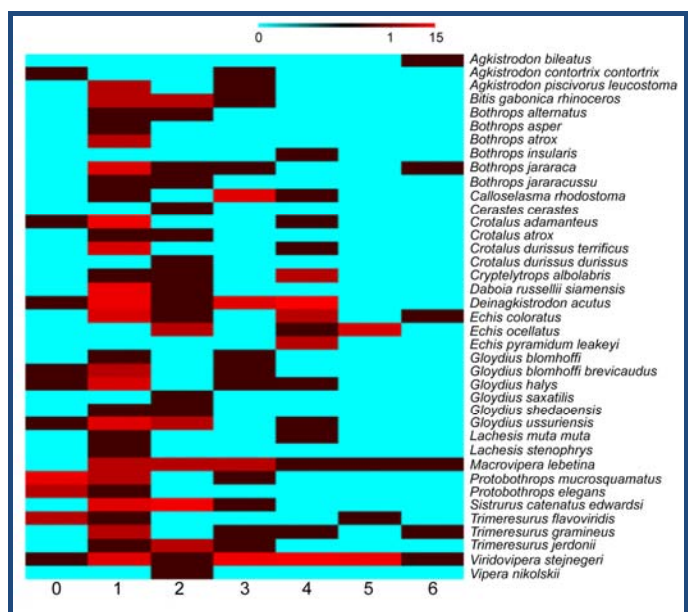
stages of the blood coagulation system, activate platelets and directly act upon fibrinogen. These include pro-coagulant enzymes such as thrombin-like enzymes which clot fibrinogen (fibrinogenolytic), factor V activators, kininogenases and platelet aggregators, and anti-coagulant enzymes such as fibrinolytic enzymes, plasminogen activators and protein C activators [10].

A detailed understanding of the components of snake venoms is important both for acquiring a more complete understanding of the pathology of envenomation and also to aid in the development of improved treatments for snakebites. Moreover, several venom enzymes, including VVSPs have proved to have potential as therapeutics for various human haemostatic disorders [11]. Despite their high sequence similarity, VVSPs differ widely in their functions. Accelerated evolution [12], exon

switching [13] and point mutations [14] have been reported to be involved in the generation of novel VVSPs and adaption to different geographical locations and available prey. Due to their physiological and medical importance, understanding of VVSP sequences, structures, functions and phylogenetic relationships represent research priorities. In this article, we report the collection of largest dataset of available VVSP sequences from public databases and literature and the detailed analysis of their sequence, structure, function and phylogenetic relationships.



**Figure 1:** Number of VVSPs with different isoelectric points in each snake. The isoelectric point of all VVSPs was predicted and clustered together for every snake based on the number of VVSPs found with varying isoelectric points. Blue colour indicates the low number (0) and the red colour indicates highest number (17) of VVSPs.



**Figure 2:** Number of VVSPs with different N-glycosylation sites in each snake. The potential N-glycosylation sites of all VVSPs were predicted and clustered together for every snake based on the number of VVSPs found with varying numbers of predicted

sites. Blue colour indicates the low number (0) and the red colour indicates highest number (15) of VVSPs.

## Methodology:

### Collection of VVSP sequences

Basic local alignment search tool (BLAST) [15] was used to search the NCBI protein database using several VVSP amino acid sequences. The NCBI and Swiss-Prot protein databases were searched with individual viper names and several keywords including serpentes, colubroidea, viperidae, viperinae and crotalinae to extract all VVSP sequences. Some published protein sequences, which had not been submitted to any of the sequence databases were collected directly from the literature. Several VVSP sequences were incomplete or contained internal deletions (longer than 10 amino acids), and were therefore excluded from our dataset. For each complete VVSP sequence the viper name, accession number, name (and alternative names if available) of the protein, experimentally determined functions and associated references were collected. Where two or more sequences had 100% sequence identity as determined by the EMBOSS pairwise alignment tool [16], duplicates were removed. Each sequence was allocated a reference code, determined using the viper's scientific name followed by up to three letters representing the protein name (e.g. CDD-SP, where CDD represent the snake name, *Crotalus durissus durissus* and SP represent the protein name, serine protease). Where, the similar first letter of two or more species occurs within a genus, the second letter of species name was used in lower case (e.g. CAd-CR).

### Sequence analysis

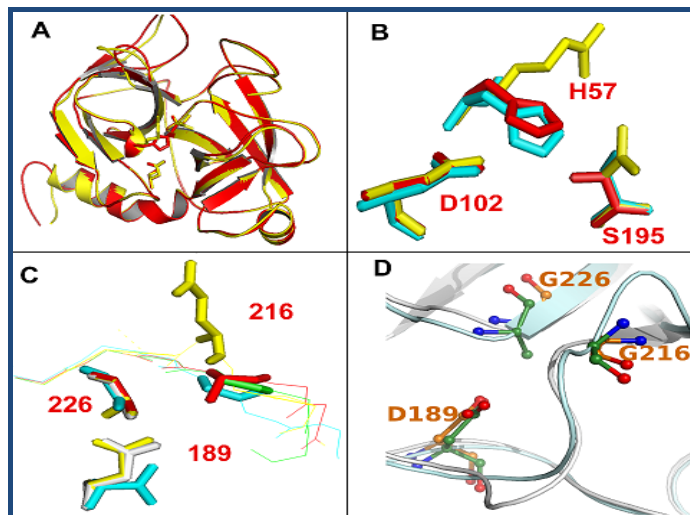
Multiple sequence alignment analysis of all the VVSP sequences was performed using the ClustalW2 multiple sequence alignment tool [17]. To aid comparison of the VVSP sequences with other serine proteases, the sequence of bovine  $\alpha$ -chymotrypsinogen (NCBI accession number: P00766) was included in the alignment. This enabled the commonly used chymotrypsinogen numbering pattern to be identified in the VVSP sequences. The alignment file was viewed using the Jalview-multiple sequence editor tool [18]. In some of the VVSP sequences (obtained from cDNA sequencing), the N-terminal signal and activation peptides were identified in comparison to previous literature and these were removed manually before the complete analysis. NetNGlyc tool was used to predict N-glycosylation sites, and compute pI/MW tool from ExPASy bioinformatics portal was used to predict the isoelectric points and molecular weight of VVSPs. The number of N-glycosylation sites and isoelectric points of VVSPs in every snake were clustered using CIMminer software [19].

### Phylogenetic analysis

To generate the phylogenetic tree, sequences were aligned using ClustalW within MEGA 4 [20] using a gap opening penalty of 10 and a gap extension penalty of 0.1 for the initial pairwise alignment and a gap opening penalty of 3 and gap extension penalty of 1.8 for the multiple alignment and the Gonnet protein weight matrix. The phylogenetic tree was generated within MEGA 4 using the neighbour-joining method and the Jones-Taylor-Thornton substitution model. The bootstrap test was performed using 2000 replications. The sequence of bovine  $\alpha$ -chymotrypsinogen was used as an outgroup.

## Structural modeling

Structural models of VVSPs were created using the IntFOLD server [21]. Models were obtained for each sequence using the structure of rat trypsin (PDB code: 1co9) as a best matched template. Models were visualised and compared with each other and the structures of bovine  $\alpha$ -chymotrypsin (PDB code: 1YPH) and rat trypsin (PDB code: 1CO9) using PyMOL molecular graphics system, version 1.3 (DeLano Scientific).



**Figure 3:** Structural models of VVSPs. Structural models of rhinoceras 2 (BG-RHIN2) and 4 (BG-RHIN4) were created using the IntFOLD server [17] using the structure of rat trypsin (PDB code: 1co9) as a template. **(A)** Schematic diagram showing the overall similarities in structure between BG-RHIN2 (yellow) and BG-RHIN4 (red). The side chain atom positions for the catalytic triad residues are included. **(B)** Detailed view of the amino acids corresponding to the catalytic triad residues in BG-RHIN2 (yellow), BG-RHIN4 (red) and chymotrypsin (PDB code: 1ypH; cyan). BG-RHIN2 has substitutions for the serine and histidine residues. **(C)** Detailed view of the main constituents of the S1 specificity pocket in BG-RHIN2 (yellow), BG-RHIN4 (red), chymotrypsin (cyan) and trypsin (PDB code: 1co9; green). In chymotrypsin these residues are: S189 at the base of the specificity pocket, with G216 and G226 at the sides. In trypsin D189 is at the base of the pocket, with G216 and G226. All images were generated using PyMOL. **(D)** Detailed view of the main constituents of the S1 specificity pocket in BG-RHIN2 (green) and protein C activator (PDB code: 2AIQ; purple).

## Results and Discussion:

### Collection of VVSP sequences

The collection of available VVSP sequences from the protein databases represented an initial step towards developing a comprehensive dataset of all VVSPs. BLAST was used to search the NCBI protein database using the amino acid sequences of several VVSPs in order to obtain all available sequences of VVSPs. In addition, the NCBI and Swiss-Prot databases were searched using list of viper names and several key words. There were several snake venom serine protease sequences from other than viperiade family members but since our main focus was on VVSPs, non-viper sequences were removed from the dataset. In total 196 complete sequences (after removing the duplicates and partial sequences) were obtained including 11 complete sequences extracted from relevant literature **Table 1 (see supplementary material)**. Of the 196 available proteins only a

small proportion (49 out of 196 or 25%) of VVSPs has been experimentally characterised. Some (9) VVSPs have been shown to have more than one function and others have not been screened for more than one particular function of interest. Some VVSPs have been designated as thrombin-like enzymes or KN (which might mean kininogenase), but no information was available regarding the basis of this classification. For several available VVSP sequences, associated published article or information regarding respective functions and other characteristics were not available. Structural information for the VVSPs is currently limited with structures solved for only 5 (2.6% in total) VVSPs known.

Our comprehensive dataset of VVSPs has been acquired using semi-automated methods and to our knowledge, this is the most complete collection to date of VVSP sequences. The absence of some sequences from any of the sequence databases makes it impossible to acquire this completely automatically. The use of different naming conventions by various authors also makes it difficult to obtain the VVSP sequences using keyword searches. The VVSP sequences along with other serine protease sequences from reptiles (other than viper snakes and lizards) grouped together in BLAST analysis, indicating them to have higher sequence similarities to each other than to non-reptile serine proteases.

### General characteristics of VVSP sequences

To analyse the various features of VVSPs, all sequences were aligned together along with bovine  $\alpha$ -chymotrypsinogen. Initial multiple sequence alignments of VVSPs demonstrated that many sequences occurred in precursor form, containing signal (18 amino acids) and activation peptides (6 amino acids) at the N-terminus. These were removed from all VVSP sequences and only the mature protein sequences were aligned for sequence analysis (Figure S1- this figure is available from authors). The bovine  $\alpha$ -chymotrypsinogen-based sequence numbering was used throughout the study to allow easier discussion with non-VVSP sequences.

### Activation of VVSPs

Most of eukaryotic proteases (e.g. digestive enzymes) are synthesised as precursors and become mature enzymes at their target sites following the cleavage of signal/activation peptides. As snake venoms are synthesised from the venom glands and the venom facilitates the immobilization and digestion of prey, they are also synthesised as inactive zymogens with signal and activation peptides [12]. These signal and activation peptides are only present in the sequences which were obtained through cDNA sequencing and were absent in any of the VVSP sequences obtained through Edman N-terminus sequencing methods from the purified proteins. A previous study [21] reported that venom proteins are stored at an acidic pH in the lumen of main secretory gland where they are maintained in their zymogen state to prevent unwanted proteolytic activity in the venom gland. During venom extraction or a bite, the venom flows through the rostral portion of accessory gland which is rich in serous secretions (contain proteolytic enzymes) that result in activation of proteolytic enzymes [22].

### Isoelectric point of VVSPs

The isoelectric point of each VVSP sequence was predicted in order to analyse the distribution of VVSPs with varying



characteristics within and among other snake venoms. The predicted isoelectric point of each VVSP varies between 4.6 and 9.4 (**Figure 1**). Depending upon the number of VVSP sequences known in each venom, the distribution of VVSPs vary. The majority of VVSPs present in all the snake venoms analysed lie between isoelectric points of 5 and 7 or 8 and 9. This analysis indicates that snakes produce a mixture of serine proteases with varying isoelectric points and these may result in alteration of their functional properties to adapt various substrates in diverse preys.

### *N-glycosylation sites in VVSPs*

Since glycosylation in two of the VVSPs (DA-ASI and DA-ASII) has been demonstrated to affect the substrate specificity of these enzymes [23], the potential N-glycosylation sites were predicted for all VVSP sequences. The number of predicted N-glycosylation sites in VVSPs varies between 0 and 6 (**Figure 2**). This analysis indicates that although some (16) of VVSPs are not glycosylated, many of the VVSPs have one (68) or more (112) potential glycosylation sites. Glycosylation of VVSPs may represent a general feature that contributes to the determination of substrate specificity similar to DA-ASI and DA-ASII. Moreover, the addition of glycosylation moieties will increase the molecular weights of native VVSPs in actual venom. The difference in molecular weights of VVSP homologs in different species is well reported in literature [14].

### *GWG motif*

The conversion of inactive zymogens into active enzymes is a key step to initiate the activity of several digestive and clotting enzymes at their target sites. This conversion is usually takes place by proteolytic cleavage at specific sites and by altering the structural fold of active enzymes. Fehlhhammer *et al.* [24] reported that a GWG motif between residues 140 and 142 in chymotrypsinogen was an important site for the conversion of a flexible zymogen structure into a fixed rigid orientation in the activated chymotrypsin. This motif is absolutely conserved in all VVSPs analysed (Figure S1), suggesting that it may play roles in the conversion of inactive to active VVSPs.

### *C-terminal extension*

The sequence alignment shows that the length of most of the mature VVSPs is between 228 and 239 amino acid residues. Compared to chymotrypsinogen, all the VVSPs lack 15 amino acids at the N-terminus of the mature proteins which indicates the removal of signal and activation peptides. While most (179) of the VVSPs contain a C-terminal extension of 7 amino acids, some (17) contain smaller or larger extension (Figure S1). This C-terminus feature is not present in chymotrypsin, trypsin, thrombin and other known mammalian serine proteases, suggesting this to be a unique feature of VVSPs. This C-terminal extension contains a cysteine residue which may form a disulphide linkage with C91 in all the VVSPs to offer additional stability for maintaining their structures as determined for VS-VPA [25].

### *Stability of VVSPs*

#### *Salt bridge*

Salt bridges are usually formed between an acidic and a basic residue to stabilise the structure. A common salt bridge in trypsin and VVSPs was reported between the residues V/I16 and D194 [25, 26]. This is formed after the activation of the

enzyme from its precursor form and is involved in stabilisation of the active site structure. All VVSPs have either valine or isoleucine at position 16 (except one VVSP, EP-SP1 which has a deletion at this position) and aspartate at position 194 (except VS-KN10 which has valine at this position) (Figure S1). Thus with the exception of these two VVSPs, all others are likely to form a salt bridge between the residues in these positions in order to increase the stability of the enzymes.

### *Disulphide linkages*

The trypsin-like serine proteases contain 10 cysteine residues which normally form 5 disulphide linkages. But in the experimentally analysed VVSPs 12 cysteine residues are found, and in the structure of VS-VPA they have been shown to form 6 disulphide linkages [25]. These are C22-C157, C42-C58, C91-C245e, C136-C201, C168-C182 and C191-C220. It is likely that similar disulphide bonds may form in all other VVSPs where these cysteine residues present at corresponding position. The disulphide linkage C91-C245e is a unique characteristic to VVSPs [25]. In VVSP sequence alignment (Figure S1), although 7 of the cysteine residues [C22, C136, C157, C168, C182, C191 and C220] are absolutely conserved in all VVSPs, 13 VVSPs have substitutions for one of the other 5 cysteine residues **Table 2 (see supplementary material)**. C42 is substituted in 4 VVSPs (CC-CR, VS-KN5, VS-KN6 and VS-KN14); C58 is substituted in one (DA-TL4); C91 is present in all VVSPs except for two (BAI-BH and GBB-SL); C245e was substituted in three VVSPs, EO-SP2, CAd-SP5 and EC-SP6, and a deletion was found in EO-SP5 at this position. C201 was substituted in two VVSPs, DA-DK and DA-TL7. In some VVSPs, the C91 has been aligned at the positions of 90 or 92, but these form disulphide linkages as normal in modelled structures (data not shown). This analysis shows that in some of the VVSPs, specific disulphide linkages cannot be formed due to the substitutions. It is entirely not clear how these substitutions and subsequent loss of corresponding disulphide linkages would affect the architecture of the VVSP structures and their function. GBB-SL which lacks C91-C245e, however, has been functionally characterised to have  $\alpha$  and  $\beta$  fibrinogenolytic activities [information obtained from the NCBI protein database].

### *Functional characteristics of VVSPs*

#### *Catalytic triad*

Trypsin-like serine proteases share a catalytic triad which includes H57, D102 and S195 (bovine  $\alpha$ -chymotrypsinogen numbering) [25]. These residues are conserved in majority of the VVSPs (Figure S1) and so they would be expected to be catalytically active. 21 VVSPs have substitutions for one or two of the catalytic triad residues. Substitutions to the catalytic triad residues have previously been observed in a small number of VVSPs [12, 14, 27, 28]. The majority of these sequences have been identified at transcript level only; prior to this study only four serine proteases (BG-RHIN2 & 3, BAI-BH and TJ-SPH) with catalytic triad substitutions have been identified within the venom of snakes [14, 28, 29]. Functional analysis of one of these proteins (TJ-SPH) has confirmed this to be functionally inactive [28], but, interestingly, another VVSP, bhalternin (BAI-BH) with substitutions at positions 57 (N) and 102 (T) has been shown to be functionally active [29]. It is therefore hard to conclude if these VVSPs would be functionally active or inactive with substitutions within the catalytic triad. The serine proteases with catalytic triad substitutions are commonly called serine

protease homologues [28]. **Table 3** (see **supplementary material**) shows the VVSPs with catalytic triad substitutions. The most frequently substituted residue within the catalytic triad is H57, which has been substituted by R in 15 sequences, by N in 3 sequences and by Q in one sequence. Ten of the sequences have substitutions for the catalytic S195; in five VVSPs this has been substituted by N, one protein has a S195P substitution, three have a S195T substitution and one has alanine at this position. In contrast, D102 is conserved in all the sequences except one, BAI-BH which has threonine at this position.

Structural models of these VVSPs show that they are generally similar to each other, consistent with the sequence similarity of the enzymes, but they differ at their catalytic site. For example, rhinocercases, BG-RHIN2 and 3 contain catalytic triad substitutions at positions 57 and 195, but show overall similarity to the predicted structures of other rhinocercases, BG-RHIN4 and 5 which contain normal catalytic triad (**Figure 3A**). They show differences at their catalytic site (**Figure 3B**). D102 is at almost identical positions in rhinocercases 2 and 4, and bovine chymotrypsin. The catalytic serine is at a very similar position in rhinocercase 4 and chymotrypsin, and the substituted N in BG-RHIN2 is also similarly located. However the substitution of the long arginine side chain instead of histidine at position 57 in BG-RHIN2 has resulted in a significant change to the orientation of the side chain in the modelled structure (**Figure 3B**). Similarly other VVSPs with different amino acid substitutions show differences in their catalytic sites.

Proteins with substitutions to the catalytic triad are present in many enzyme families; indeed it has been estimated that up to 15% of the members of all encoded enzyme families may have lost their catalytic activity [30]. Siigur *et al.* [12] suggested that the critical substitutions in the VVSP, ML-P2 from the venom of *Macrovipera lebetina* could be generated via trans-splicing of the primary gene transcript, exon-shuffling or unequal crossing over at the genome level. We have however, shown previously that these substitutions may have occurred at multiple levels [14]. Although one of the VVSPs with a catalytic triad substitution was proved to be functionally active [29], another was shown to be inactive [28]. No further VVSP of this nature has been functionally characterised. So it is not entirely clear if these proteins are functionally active in the venom.

In many cases in other biological systems, inactive homologues are believed to have acquired alternative functions, such as competing with and antagonising the active proteases, or otherwise regulating their function. Within invertebrates, serine protease homologues have been shown to be involved in various defence responses [31]. It has however, been suggested that some invertebrate serine protease homologues are unlikely to bind peptide substrates by a canonical protease-like mechanism, though other potential protein binding sites have been suggested [32]. Within snake venom, catalytically inactive phospholipase A<sub>2</sub> such as myotoxins II and IV from the venom of *Bothrops asper* are known to act as toxins and are thought to bind to their target membrane substrates in order to reduce their permeability control and cause subsequent necrosis [33]. Corey *et al.* [34] reported that trypsin with a H57R substitution had 190000 fold lower activities than the wild type enzyme at pH 8.0 but had 38 fold higher activities than the wild type at pH

10.1. Clearly, experiments to determine the function of the serine protease homologues within snake venom need to be performed. It is possible that they may affect the physiology of victims or prey by binding irreversibly to substrates involved in blood coagulation and preventing their normal function.

### Specificity of VVSPs

The primary specificity pocket of serine proteases (also called as S1 binding site) accommodates the side chain of the P1 residue (the residue preceding the scissile bond) of substrates and confer specificity towards substrates. For example, chymotrypsin contains S189, G216 and G226 and is specific for the large hydrophobic side chains (such as F, Y and W) [35]. Trypsin and thrombin contain D189, G216 and G226 and are specific for positively charged amino acids such as R and K [35]. Elastase contains S189, V216 and T226 and is specific for small hydrophobic side chains like alanine [35]. In addition to the catalytic triad residues, there are also differences found in the residues present in the primary specificity pockets of several VVSPs (Figure S1). In total 161 VVSPs have D189, G216 and G/A226 and so they may confer specificity towards either or both positively charged amino acids such as R and K at P1 position of substrates similar to trypsin and thrombin.

The remaining 35 VVSPs, however, have different combinations of specificity pocket residues **Table 4** (see **supplementary material**). Although seven VVSPs contain aspartate at position 189, they have substitutions in the other two sites and thus, they might restrict entry for substrates. For example, in the venom of *Bitis gabonica rhinoceros*, BG-RHIN2 and 3 have D189, which might indicate specificity for basic residues but have E instead of G at position 216 and A instead of G at position 226 (**Figure 3C & 3D**). These substitutions are likely to restrict access to the specificity pocket, thus the binding specificity of BG-RHIN2 and 3 is not clear. In contrast, two other VVSPs (BJa-HP3 and AP-PA), contain D189-R216-G226 in their specificity pocket. So the presence of arginine at position 216 may create a basic environment in the pocket. The functions of these VVSPs are further critical as they contain D at 189 and R at 216. A further two VVSPs (TG-SP3 and VS-KN8) contain D189-G216-N/V226 in their specificity pocket. Although they can confer specificity towards basic amino acids in their substrates, the presence of N or V at position 226 will have influence in their specificity.

The remaining 28 VVSPs contain substitution mainly for position 189 **Table 4** (see **supplementary material**). Thus, these enzymes may not confer specificity for arginine at P1 position of the substrates. For example, 13 VVSPs have G189-G216-G226 in their specificity pocket. This is an unusual substitution in trypsin-like serine proteases, which is similar to the human kallikrein KLK9 whose specificity is unknown [36]. The substitution might be expected to increase the size of the specificity pocket; *M. lebetina* α and β-fibrinogenases (ML-AF and ML-BF) also have G189 [37] and Siigur *et al.* [38] have reported that ML-AF hydrolysed the Y16 - L17, F24 - F25 and F25 - Y26 bonds of the insulin B chain, suggesting that enzymes with G189 can cleave substrates with large hydrophobic residues at the P1 position. Some VVSPs such as BG-RHIN4 and 5 also have a G to A substitution at position 216 which may narrow the specificity pocket slightly. There are also other

combinations of residues present in small number of VVSPs whose functions are unknown.

Comparison of the positions and orientations of residues in the S1 specificity pockets of BG-RHIN2 and 4 with those of chymotrypsin and trypsin suggest that, the bottom of the specificity pocket is very similar in BG-RHIN2 and trypsin, while the glycine in BG-RHIN4 is uncharged and protrudes even less into the pocket (Figure 3C). At position 226, the glycines and alanines in all enzymes are very similarly located. In contrast, the substitution of the large negatively charged glutamic acid at position 216 clearly has a significant effect on the S1 pocket in BG-RHIN2 when the modelled structure was compared to model of BG-RHIN4 (Figure 3C). When BG-RHIN2 model was superimposed on the determined structure of protein C activator (AC-AN) [39], G217 of BG-RHIN2 is situated at the same position as G216 of AC-AN (Figure 3D). So it is not entirely clear if negatively charged glutamate or glycine would occupy this position in the real structure of BG-RHIN2. However the primary specificity pocket is not the sole determinant of specificity. For serine proteases involved in coagulation, the importance of additional regions of the structure e.g. exosites in recognition of substrates is becoming increasingly recognised [40]. Thus the precise specificity of these enzymes can only be determined experimentally. The potential role of exosites in binding to substrates also strengthens our suggestion above that the serine proteases homologues which lack the catalytic triad may still be capable of interfering with the coagulation cascade through exosite-mediated interactions.

### Oxyanion hole

A further interesting feature of VVSP sequences is that G193, which is generally highly conserved in serine proteases and is involved in the formation of oxyanion hole and in inhibitor binding [41], is substituted in 57 sequences: R in 25, A in 7, V in 6, F in 6, Q in 4, H in 3, C in 2, Y in 2, S in 1 and T in 1 VVSP (Figure S1). As for VS-VPA, a venom plasminogen activator from *Trimeresurus stejnegeri*, which has F at position 193 [41], bulky residues substituted for G193 may reduce the sensitivity of the proteins to inhibitors such as bovine pancreatic trypsin inhibitor as well as reduce their interaction with substrates. Similarly other residues including charged residues such as arginine and glutamine may influence the formation of oxyanion hole and inhibitor or substrate binding.

### RGD motif

The arginine-glycine-aspartate (RGD) motif acts as a cell adhesion signal and is present in cellular signal transduction proteins. It has also been reported that proteins containing RGD motifs may induce apoptosis [42]. Lee *et al.* [43] identified a RGD motif in brevinase (GB-BV) and suggested that it might be involved in cellular recognition and in inhibition of platelet aggregation through binding to integrin  $\alpha_{IIb}\beta_3$ . In our analysis we have identified the RGD motif at the same position as GB-BV (between 192 and 194), in four other VVSPs: GB-HY, GH-H, GU-TC and GSh-SH (Figure S1). VVSPs containing the RGD motif could be involved in inducing cell death in order to cause destruction of cells in the host system. They may also inhibit platelet aggregation. As RGD motifs were found only in VVSPs from venoms of the *Gloydius* genus so far, this may be a unique feature of these snake venoms.

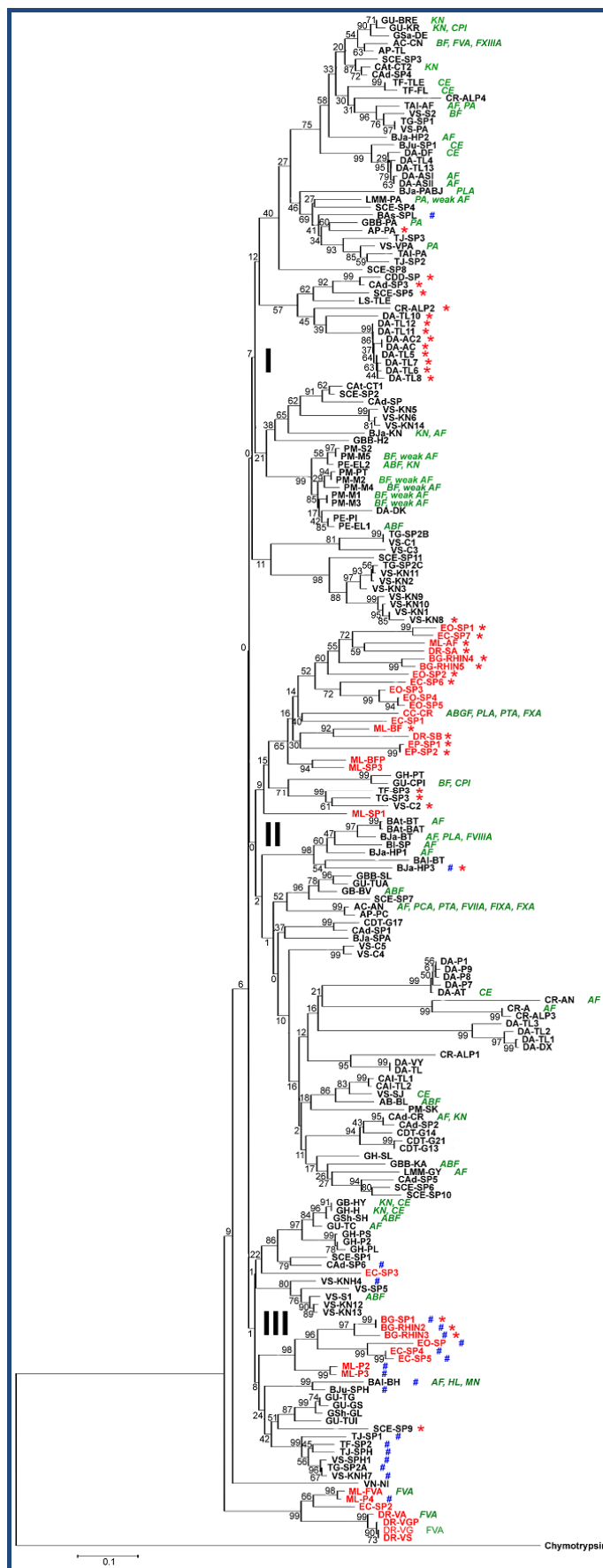
### Phylogenetic analysis of VVSPs

Phylogenetic analysis was performed to analyse the evolutionary and functional relationships of VVSP sequences. The phylogenetic tree generated for 196 VVSP sequences is shown in (Figure 4). The sequence of bovine  $\alpha$ -chymotrypsinogen was displayed as an out-group in the tree. Wang *et al.* [44] generated a phylogenetic tree with 41 VVSP sequences using the Phylip-neighbor method. From their tree, three categories of VVSPs were identified, which they named kininogenases (KN), clotting enzymes (CL) and plasminogen activators (PA). These names described the known functions of some members of each of the categories. They also suggested that the KN category might be divided into 2 subcategories, with one representing GBH and another representing PMM1-PMM5 and that this division might become more apparent if more sequences, particularly from true vipers were added. This classification was not apparent in our analysis with 196 sequences, which cluster together in 3 major groups (named I, II and III).

The sequences from true vipers (viperinae subfamily) are mostly clustered together in group II and III (shown as red letters in Figure 4). Moreover, the VVSPs with catalytic triad substitutions (shown as # in the Figure 4) cluster together in group III except two VVSPs which cluster in I and II. Similarly, the sequences with substituted primary specificity pocket are clustered in groups I and II, with an exception of one VVSP which cluster in group III (shown as \* in the Figure 4). This suggests that the VVSPs from true vipers differ evolutionarily from the pit vipers. Moreover, the evolution of VVSPs with substitutions in the catalytic triad and specificity pocket is not a random process in any particular snake; instead it might be a controlled mechanism to develop novel VVSPs with differing characteristic features.

The experimentally determined functions of some VVSPs are indicated (as green letters in Figure 4) in the tree. Since majority of the functionally characterised VVSPs are capable of performing more than one function, the classification of VVSPs based on their function is difficult. Moreover, several VVSPs have been characterised only for the functions of interest and thus, they were not screened for other functions which they may perform. This analysis indicates that the classification of VVSPs based on some of the functionally characterised VVSPs as previously reported [44], may not appropriate with all the available VVSPs. For example, the VVSPs with kininogenase activities are clustered in all three groups in the tree. Similarly the coagulant enzymes such as  $\alpha$ -fibrinogenase (AF),  $\beta$ -fibrinogenase (BF) and  $\alpha\beta$ -fibrinogenase (ABF) are present randomly in all clusters. Some enzymes such as plasminogen or factor V activators are clustered together in a particular group in the tree, although only a very small number of VVSPs with those functions are identified to date. Moreover, the number of VVSPs performing additional functions such as platelet aggregator, factor VIII/X/XIII activator, an enzyme increases capillary permeability and albuminolytic/ myonecrotic/ hemolytic enzymes are limited and thus it is not apparent to predict the functions of other VVSPs which cluster together with these enzymes.





**Figure 4:** Phylogenetic tree of VVSPs. 196 VVSP sequences from were included for the generation of a phylogenetic tree together with bovine  $\alpha$ -chymotrypsinogen (NCBI accession ISSN 0973-2063 (online) 0973-8894 (print) Bioinformatics 8(16): 763-772 (2012)

number: P00766) as an out-group. The alignment was generated using ClustalW within MEGA 4 using a gap opening penalty of 10 and a gap extension penalty of 0.1 for the initial pairwise alignment, gap opening penalty of 3 and gap extension penalty of 1.8 for the multiple alignment and the Gonnet protein weight matrix. The phylogenetic tree was generated from this within MEGA 4 using the neighbour-joining method and the Jones-Taylor-Thornton substitution model. The bootstrap test was done using 2000 replications. The VVSPs from true vipers (viperinae sub family) are shown in red. # and\* indicate the sequences with catalytic triad and specificity pocket substitutions respectively. The experimentally determined functions of VVSPs are shown in green letters with short codes; AF-  $\alpha$  fibrinogenase, BF-  $\beta$  fibrinogenase, ABF-  $\alpha\beta$  fibrinogenase, KN- kininogenase, CPI- capillary permeability increasing enzyme, FVA- factor V activator, FXIIIa- factor XIII activator, CE- clotting enzyme, PA- plasminogen activator, PLA- platelet activator, ABGF-  $\alpha\beta\gamma$  fibrinogenase, PTA- prothrombin activator, FXA- factor X activator, FVIIIa- factor VIII activator, FVIIa- factor VII activator, PCA- protein C activator, FIXA- factor IX activator and HL- haemolytic enzyme

### Conclusion:

To our knowledge this study represents the most complete information about VVSPs available to date and this will serve as a dataset for further analysis in these enzymes. VVSPs share several sequence features such as catalytic triad and disulphide linkages with other common serine proteases for example, chymotrypsin, trypsin and thrombin. Some features, such as the C-terminal extension of 7 residues and an additional disulphide linkage, are unique to VVSPs. However, within the VVSP family there are many sequence variations which might be responsible for the variation in specificity of these enzymes. The phylogenetic analysis of VVSPs shows three clusters of sequences and the classification of these enzymes based on only a few functionally characterised enzymes may not be appropriate with a larger dataset. Our analysis showed that the naming or prediction of function for each cluster may not be possible as previously suggested when we include more sequences as most of them are of unknown functions and some of them are only partially characterised. A greater level of functional characterisation is needed to obtain more complete functionally known VVSPs for better understanding of their sequence, function and evolutionary relationships. Posttranslational modification predictions showed that most of the VVSPs were predicted to be N-glycosylated. The functions of these modifications in VVSPs are yet to be clearly identified but, in several other proteins it has been shown that glycosylation affects the function of the protein. Structural analysis of VVSPs has helped to identify the functionally important residues; catalytic triad, primary specificity pocket and disulphide linkages. However only 5 determined structures of VVSPs are available to date and thus more crystallisation work should be performed to obtain more structures. The dataset presented here with complete information of all known VVSPs to date will form a basis for further research to explore more about the sequence, structure, function and phylogenetic relationships of these enzymes in order to develop improved therapeutics for the treatment of snakebites.

### References:

[1] White J, *Ther Drug Monit.* 2000 **22**: 65 [PMID: 10688262]

- [2] Chippaux JP, *Bull World Health Organ.* 1998 **76**: 515 [PMID: 9868843]
- [3] Kasturiratne A *et al.* *PLoS Med.* 2008 **5**: e218 [PMID: 18986210]
- [4] Vaiyapuri S *et al.* *PLoS Negl Trop Dis.* 2010 **4**: e796 [PMID: 20706583]
- [5] Vaiyapuri S *et al.* *PLoS One.* 2010 **5**: e9687 [PMID: 20300193]
- [6] Vaiyapuri S *et al.* *J Biol Chem.* 2012 **287**: 26235 [PMID: 22689571]
- [7] Aird SD, *Toxicol.* 2002 **40**: 335 [PMID: 11738231]
- [8] Serrano SM & Maroun RC, *Toxicol.* 2005 **45**: 1115 [PMID: 15922778]
- [9] Meier J & Stocker K, *Crit Rev Toxicol.* 1991 **21**: 171 [PMID: 2039593]
- [10] Markland FS, *Toxicol.* 1998 **36**: 1749 [PMID: 9839663]
- [11] Atkinson RP, *Cerebrovasc Dis.* 1998 **1**: 23 [PMID: 9547025]
- [12] Siigur E *et al.* *Gene.* 2001 **263**: 199 [PMID: 11223258]
- [13] Doley R *et al.* *BMC Evol Biol.* 2009 **9**: 146 [PMID: 19563684]
- [14] Vaiyapuri S *et al.* *PLoS One.* 2011 **6**: e21532 [PMID: 21731776]
- [15] Altschul SF *et al.* *J Mol Biol.* 1990 **215**: 403 [PMID: 2231712]
- [16] Rice P *et al.* *Trends Genet.* 2000 **16**: 276 [PMID: 10827456]
- [17] Larkin MA *et al.* *Bioinformatics.* 2007 **23**: 2947 [PMID: 17846036]
- [18] Waterhouse AM *et al.* *Bioinformatics.* 2009 **25**: 1189 [PMID: 19151095]
- [19] Scherf U *et al.* *Nat Genet.* 2000 **24**: 236 [PMID: 10700175]
- [20] Kumar S *et al.* *Brief Bioinform.* 2008 **9**: 299 [PMID: 18417537]
- [21] Roche DB *et al.* *Nucleic Acids Res.* 2011 **39**: W171 [PMID: 21459847]
- [22] Mackessy SP & Baxter LM, *Zoologischer Anzeiger.* 2006 **245**: 147
- [23] Zhu Z *et al.* *Acta Crystallogr D Biol Crystallogr.* 2003 **59**: 547 [PMID: 12595722]
- [24] Fehlhammer H *et al.* *J Mol Biol.* 1977 **111**: 415 [PMID: 864704]
- [25] Parry MA *et al.* *Structure.* 1998 **6**: 1195 [PMID: 9753698]
- [26] Massova I *et al.* *Bioorganic & Medicinal Chemistry Letters.* 1997 **7**: 3139
- [27] Francischetti IM *et al.* *Gene.* 2004 **337**: 55 [PMID: 15276202]
- [28] Wu J *et al.* *Toxicol.* 2008 **52**: 277 [PMID: 18590752]
- [29] Costa Jde O *et al.* *Toxicol.* 2010 **55**: 1365 [PMID: 20184912]
- [30] Pils B & Schultz J, *J Mol Biol.* 2004 **340**: 399 [PMID: 15210342]
- [31] Ross J *et al.* *Gene.* 2003 **304**: 117 [PMID: 12568721]
- [32] Fischer K *et al.* *J Mol Biol.* 2009 **390**: 635 [PMID: 19427318]
- [33] Angulo Y & Lomonte B, *Toxicol.* 2009 **54**: 949 [PMID: 19111755]
- [34] Corey DR & Craik CS, *J Am Chem Soc.* 1992 **114**: 1784.
- [35] Czapinska H & Otlewski J, *Eur J Biochem.* 1999 **260**: 571 [PMID: 10102985]
- [36] LeBeau AM *et al.* *Biochemistry.* 2009 **48**: 3490 [PMID: 19281249]
- [37] Siigur E *et al.* *Thromb Haemost.* 2003 **89**: 826 [PMID: 12719779]
- [38] Siigur J *et al.* *Procedia Chemistry.* 2010
- [39] Murakami MT & Arni RK, *J Biol Chem.* 2005 **280**: 39309 [PMID: 16162508]
- [40] Krishnaswamy S, *J Thromb Haemost.* 2005 **3**: 54 [PMID: 15634266]
- [41] Braud S *et al.* *J Biol Chem.* 2000 **275**: 1823 [PMID: 10636881]
- [42] Buckley CD *et al.* *Nature.* 1999 **397**: 534 [PMID: 10028971]
- [43] Lee JW *et al.* *Biochem Biophys Res Commun.* 1999 **260**: 665 [PMID: 10403823]
- [44] Wang YM *et al.* *Biochem J.* 2001 **354**: 161 [PMID: 11171091]

Edited by P Kanguane

Citation: Vaiyapuri *et al.* *Bioinformation* 8(16): 763-772 (2012)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited



## Supplementary material:

**Table 1:** The dataset of VVSPs used in the analysis. In total 185 VVSP sequences were collected from the protein sequence databases and 11 sequences were obtained from relevant literature (indicated as 'not reported' under accession number category). The name of snake, NCBI accession number and the short reference code for each sequence is given in the table

Viper Species	Accession number	Reference code	Viper Species	Accession number	Reference code
<i>Agkistrodon bileatus</i>	Q9PSN3	AB-BL		AD147557	EP-SP2
<i>Agkistrodon contortrix contortrix</i>	P82981	AC-CN	<i>Gloydius blomhoffi</i>	P81176	GB-HY
	P09872	AC-AN		Q9PT51	GB-BV
<i>Agkistrodon piscivorus leucostoma</i>	ADP88562	AP-PC	<i>Gloydius blomhoffi brevicaudus</i>	AAD01624	GBB-PA
	ADP88561	AP-PA		Q9YGY9	GBB-H2
	ADW77220	AP-TL		P85109	GBB-KA
<i>Bitis gabonica rhinoceros</i>	AAR24534	BG-SP1	<i>Gloydius halys</i>	Q9PTL3	GBB-SL
	CBM40645	BG-RHIN2		Q9YGI6	GH-P2
	CBM40646	BG-RHIN3		O93421	GH-PS
	CBM40647	BG-RHIN4	<i>Gloydius halys</i>	Q802F0	GH-PT
	CBM40648	BG-RHIN5		Q9YGY2	GH-PL
<i>Bothrops alternatus</i>	Q61WF1	BAI-BT		AAC61838	GH-SL
	P0CG03	BAI-BH		P81176	GH-H
<i>Bothrops asper</i>	Q072L6	BA-S-SPL	<i>Gloydius saxatilis</i>	Q75ZE1	GSa-DE
<i>Bothrops atrox</i>	P04971	BAI-BT	<i>Gloydius shedaensis</i>	Q615L0	GSh-SH
	AAA48553	BAI-BAT		P0C5B4	GSh-GL
<i>Bothrops insularis</i>	Q8QC86	BI-SP	<i>Gloydius ussuriensis</i>	Q75ZE2	GU-BRE
<i>Bothrops jararaca</i>	Q5W959	Bja-IHP2		Q8UUJ2	GU-TUI
	P81824	Bja-PABJ		Q91053	GU-TC
	O13069	Bja-KN		Q8UUJ1	GU-TUA
	Q9PTU8	Bja-SPA		O42207	GU-CPI
	P81661	Bja-BT		Q8UVX1	GU-TG
	Q5W960	Bja-IHP1		P86171	GU-KR
	Q5W958	Bja-HP3		Not reported	GU-GS
<i>Bothrops jararacussu</i>	Q2PQJ3	Bju-SP1	<i>Lachesis muta muta</i>	Q27J47	LMM-PA
	Q71229	Bju-SPH		P33589	LMM-GY
<i>Calloselasma rhodostoma</i>	P26324	CR-AN	<i>Lachesis stenophrys</i>	Q072L7	LS-TLE
	P47797	CR-A	<i>Macrovipera lebetina</i>	Q9PT41	ML-FVA
	CAA01526	CR-ALP4		Q9PT40	ML-P2
	CAA01525	CR-ALP1		Q8JH85	ML-AF
	CAA01524	CR-ALP2		Q8JH62	ML-BF
	CAA01523	CR-ALP3		Not reported	ML-P3
<i>Cerastes cerastes</i>	Q7SYF1	CC-CR		Not reported	ML-P4
<i>Crotalus adamanteus</i>	Q8UUK2	CAd-SP		ADN04917	ML-BFP
	Not reported	CAd-CR		ADN04916	ML-SP1
	AEJ32000	CAd-SP6	<i>Protobothrops mucrosquamatus</i>	ADN04918	ML-SP3
	AEJ31999	CAd-SP5		Q9DG84	PM-S2
	AEJ31998	CAd-SP4		Q91511	PM-M5
	AEJ31997	CAd-SP3		Q91507	PM-M1
	AEJ31996	CAd-SP2		Q91510	PM-M4
	AEJ31995	CAd-SP1		AAB01070	PM-PT
<i>Crotalus atrox</i>	Q8QHK2	CAI-CT2		Q9DG83	PM-SK
	Q8QHK3	CAI-CT1		Q91509	PM-M3
<i>Crotalus durissus terrificus</i>	Q58G94	CDT-G21	<i>Sistrurus catenatus edwardsi</i>	Q91508	PM-M2
	ABY65929	CDT-G13		ABG26969	SCE-SP3
	ABY65930	CDT-G14		ABG26970	SCE-SP4
	ABY65931	CDT-G17		ABG26974	SCE-SP8
<i>Crotalus durissus durissus</i>	Q2QA04	CDD-SP		ABG26975	SCE-SP9
<i>Cryptelytrops albolabris</i>	Not reported	CAI-AF		ABG26967	SCE-SP1
	Not reported	CAI-PA		ABG26977	SCE-SP11
	ABS12074	CAI-TL1		ABG26968	SCE-SP2
	ABS12075	CAI-TL2		ABG26972	SCE-SP6
<i>Daboia russelli siamensis</i>	P18965	DR-VG		ABG26976	SCE-SP10
	P18964	DR-VA		ABG26973	SCE-SP7
	ADP88560	DR-SB		ABG26971	SCE-SP5
	ADP88559	DR-SA	<i>Protobothrops elegans</i>	Not reported	PE-PI
	ADP88558	DR-VGP		P84787	PE-EL2
	3SBK_A	DR-VS		P84788	PE-EL1
<i>Deinagkistrodon acutus</i>	AAW56608	DA-TL1	<i>Trimeresurus flavoviridis</i>	P05620	TF-TLE
	AAW56633	DA-TL2		1405260A	TF-FL
	AAW56609	DA-TL3		O13057	TF-SP2
	ABK91854	DA-TL4		O13058	TF-SP3
	ABK91856	DA-TL5	<i>Trimeresurus gramineus</i>	O13059	TG-SP1
	ABK97625	DA-TL6		O13062	TG-SP2C
	ABK97628	DA-TL7		O13063	TG-SP3
	ABK91852	DA-TL8		O13060	TG-SP2A
	AAK52506	DA-TL10		O13061	TG-SP2B
	ABK97627	DA-TL11	<i>Trimeresurus jerdonii</i>	Q9DF66	TJ-SP3
	ABK91853	DA-TL12		Q9DF67	TJ-SP2
	ABK97626	DA-TL13		Q9DF68	TJ-SP1
	1OP2_A	DA-ASII		B0ZT25	TJ-SPH
	1OP0_A	DA-ASI	<i>Viridovipera stejneri</i>	Q8AY79	VS-S2
	AAD19350	DA-DF		Q71QH7	VS-PA
	Q918X0	DA-DK		1BQY_A	VS-VPA
	Q918W9	DA-DX		Q8AY82	VS-SPH1
	CAC00530	DA-AC2		Q71QJ1	VS-KN9
	CAB46431	DA-AC		Q71QI8	VS-KN10
	Q918X2	DA-AT		Q71QJ3	VS-KN1
	ACF17164	DA-VY		Q71QH5	VS-KN8
	Not reported	DA-P1		Q71QI5	VS-KN3
	Not reported	DA-P7		Q71QI7	VS-KN11
	Not reported	DA-P8		Q71QI4	VS-KN5
	Not reported	DA-P9		Q71QJ4	VS-KNH4
	ABV89584	DA-TL		Q71QI1	VS-KN12

<i>Echis coloratus</i>	ADI47576	EC-SP1	Q71QH6	VS-KN13
	ADI47574	EC-SP2	Q8AY80	VS-S1
	ADI47566	EC-SP3	Q8AY78	VS-SP5
	ADI47562	EC-SP4	Q8AY81	VS-SJ
	ADI47559	EC-SP5	AAQ02901	VS-C3
	ADI47552	EC-SP6	AAQ02898	VS-C1
	ADI47549	EC-SP7	Q71QI3	VS-C5
<i>Echis ocellatus</i>	ADE45141	EO-SP	AAQ02909	VS-C4
	ADC52861	EO-SP1	Q71QI2	VS-C2
	ADE45139	EO-SP2	Q71QI0	VS-KNH7
	ADE45140	EO-SP3	Q71QJ0	VS-KN2
	CAQ72889	EO-SP4	Q71QJ2	VS-KN6
	ADI47555	EO-SP5	Q71QH9	VS-KN14
	ADI47558	EP-SP1	CBW30778	VN-NI
<i>Echis pyramidum leakeyi</i>		<i>Vipera nikolskii</i>		

**Table 2:** VVSPs with substitutions for cysteine residues involved in the formation of disulphide linkages. The VVSPs and their substituted residue (blue) in the appropriate positions are shown in table. The positions of two corresponding cysteine residues which form disulphide linkage are placed next to each other to indicate the missing disulphide linkage due to substitution. VVSPs normally contain 6 disulphide linkages; C22-C157, C42-C58, C91-C245e, C136-C201, C168-C182 and C191-C220. X indicates the substituted residue and - indicates a deletion.

Name of VVSP	22	157	42	58	91	245e	136	201	168	182	191	220
BAI-BH	C	C	C	C	X	C	C	C	C	C	C	C
CAd-SP5	C	C	C	C	C	X	C	C	C	C	C	C
CC-CR	C	C	X	C	C	C	C	C	C	C	C	C
DA-DK	C	C	C	C	C	C	C	X	C	C	C	C
DA-TL4	C	C	C	X	C	C	C	C	C	C	C	C
DA-TL7	C	C	C	C	C	C	C	X	C	C	C	C
EO-SP2	C	C	C	C	C	X	C	C	C	C	C	C
EO-SP5	C	C	C	C	C	-	C	C	C	C	C	C
EO-SP6	C	C	C	C	C	X	C	C	C	C	C	C
GBB-SL	C	C	C	C	X	C	C	C	C	C	C	C
VS-KN5	C	C	X	C	C	C	C	C	C	C	C	C
VS-KN6	C	C	X	C	C	C	C	C	C	C	C	C
VS-KN14	C	C	X	C	C	C	C	C	C	C	C	C

**Table 3:** VVSPs with substituted catalytic triad residues. VVSPs with various substitutions (blue) for their catalytic triad residues are shown in table. Normally VVSPs contain H57, D102 and S195 in their catalytic triad.

Name of VVSP	57	102	195	Name of VVSP	57	102	195
BGSP1	R	D	N	ML-P2	R	D	N
BG-RHIN2	R	D	N	ML-P3	H	D	N
BG-RHIN3	R	D	N	ML-P4	R	D	S
BAI-BH	N	T	S	TF-SP2	R	D	S
Bas-SPL	N	D	S	TG-SP2A	R	D	S
Bja-HP3	Q	D	S	TJ-SP1	R	D	P
Bju-SPH	N	D	S	TJ-SPH	R	D	S
CAd-SP6	H	D	A	VS-SPH1	R	D	S
EC-SP4	R	D	T	VS-KNH4	R	D	S
EC-SP5	R	D	T	VS-KNH7	R	D	S
EO-SP	R	D	T				

**Table 4:** VVSPs with substituted primary specificity pocket residues. VVSPs with various substitutions (blue) for their specificity pocket residues are shown in table. Normally VVSPs contain D189, G216 and G/A226 in their specificity pocket.

Name of VVSP	189	216	226	Name of VVSP	189	216	226
AP-PA	D	R	G	DA-TL12	G	G	G
BG-SP1	D	E	A	DR-SA	G	G	S
BG-RHIN2	D	E	A	DR-SB	G	G	V
BG-RHIN3	D	E	A	EC-SP6	G	G	G
BG-RHIN4	G	A	G	EC-SP7	G	V	G
BG-RHIN5	G	A	G	EP-SP1	G	G	V
Bja-HP3	D	R	G	EP-SP2	G	G	V
CAd-SP3	G	G	G	EO-SP1	G	V	G
CR-ALP2	S	G	A	EO-SP2	G	S	G
CDD-SP	G	G	E	ML-AF	G	G	G
DA-AC	G	G	G	ML-BF	G	G	V
DA-AC2	G	G	G	SCE-SP9	S	G	S
DA-TL5	G	G	G	SCE-SP5	S	G	G
DA-TL6	G	G	G	TF-SP3	G	G	G
DA-TL7	G	G	G	TG-SP3	D	G	N
DA-TL8	G	G	G	VS-C2	G	T	A
DA-TL10	G	G	G	VS-KN8	D	G	V
DA-TL11	G	G	G				