The evolution and structure of snake venom phosphodiesterase (svPDE) highlight its importance in venom actions

Cheng-Tsung Pan[1,2], Chien-Chu Lin[1], I-Jin Lin[1], Kun-Yi Chien[3], Yeong-Shin Lin[4]*, Hsiao-Han Chang[1,5,6]*, Wen-Guey Wu[1,5]*

[1] Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, Taiwan

[2] Department of Biostatistics, University of Oslo, Oslo, Norway

[3] Graduate Institute of Biomedical Sciences, Department of Biochemistry and Molecular Biology, School of Medicine, Chang Gung University, Taiwan

[4] Institute of Bioinformatics and Systems Biology, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

[5] Department of Life Science, National Tsing Hua University, Hsinchu, Taiwan

[6] Institute of Molecular and Cellular Biology, National Tsing Hua University, Hsinchu, Taiwan

*Co-corresponding authors: Yeong-Shin Lin, Hsiao-Han Chang, Wen-Guey Wu

Email: YSL yslin@nycu.edu.tw

HHC hhchang@life.nthu.edu.tw

WGW wgwu@life.nthu.edu.tw

Author Contributions: CTP conceived and planned the study and analyzed the data. CCL solved the X-ray crystal structures and CCL and IJL performed the biochemical binding experiments. KYC carried out MASS for protein and glycan ID. WGW guided the snake venom project and toxin protein group. HHC and YSL advised the bioinformatics analysis. CTP, CCL and IJL wrote the manuscript. WGW, HHC and YSL provided comments and revised the manuscript. All authors have read and approved this manuscript.

Classification: Evolution

Keywords: phosphodiesterase, snake venom, alternative splicing, mobile elements, ectonucleotide pyrophosphatase/phosphodiesterase

This PDF file includes:

Main Text

Figures 1 to 3

Abstract

For decades, studies of snake venoms focused on the venom-ome-specific toxins (VSTs). VSTs are dominant soluble proteins believed to contribute to the main venomous effects and emerged into gene clusters for fast adaptation and diversification of snake venoms. However, the conserved minor venom components, such as snake venom phosphodiesterase (svPDE), remain largely unexplored. Here, we focus on svPDE by genomic and transcriptomic analysis across snake clades and demonstrate that soluble svPDE is co-opted from the ancestral membrane-attached ENPP3 (ectonucleotide pyrophosphatase/phosphodiesterase 3) gene by replacing the original 5' exon with the exon encoding a signal peptide. Notably, the exons, promoters and transcription/translation starts have been replaced multiple times during snake evolution, suggesting the evolutionary necessity of svPDE. The structural and biochemical analyses also show that svPDE shares the similar functions with ENPP family, suggesting its perturbation to the purinergic signaling and insulin transduction in venomous effects.

Significance Statement

We provided a case of the evolutionary co-option strategy in which the secretory svPDE, one of the minor venom components, is generated from the ancestral membrane-anchored ENPP3 gene by using an alternative 5' exon. The first exon of the svPDE transcript encodes a signal peptide instead of the transmembrane domain of the ENPP3. Multiple replacement events of genomic elements during snake evolution maintain the expression of the svPDE transcripts, suggesting the crucial function of svPDE. Also, the structures and biochemical analyses indicate that svPDE reuse the original functions of ENPP3 in the bitten tissues and may interfere with normal cell signaling. Together, our study reveals the evolution of svPDE and suggests the importance of minor venom components.

Main Text

Introduction

Snakebite envenoming is a neglected tropical disease that leads to over 100,000 deaths worldwide annually (Gutierrez et al., 2017) and its mortality may be even higher than malaria in some regions (Stock, Massougbodji, Alagon, & Chippaux, 2007). For decades, scientists have been dedicated to discovering the components of snake venoms. The diversity and species-specificity of snake venoms make the first aid and treatments difficult for snakebites. Traditionally, the dominant proteins in the venoms have been identified by proteomic approaches as the major components of venoms. Recent advances in whole-genome sequencing of venomous snakes have prompted many studies to consider the origin and diversification of such dominant proteins , which have been named as the venom-ome-specific toxins (VSTs) (Suryamohan et al., 2020) and have regarded as the main contributors of toxic effects.

Most VSTs have evolved from ancestral genes and been duplicated into large gene clusters in several snake clades. For example, the gene clusters of snake venom metalloproteases (SVMP) in Viperidae and three-finger toxins (3FTx) in Elapidae have been revealed to evolve from the ancestral genes of a membrane-anchored protein (named ADAM28 in other species) and a membrane GPI-anchored protein (named Ly6 in other species), respectively (Almeida et al., 2021; Giorgianni et al., 2020; Margres et al., 2021; Rao et al., 2022). Alternative mRNA splicing at the 3' end has been proposed to account for the adaptation of encoding ancestral membrane-anchored ADAM28 proteins into a soluble protein in Viperidae by losing the C-terminal transmembrane domain. This ancestral gene has evolved into the SVMP gene cluster with venom-specific expression in Viperidae followed by multiple gene duplication events. (Giorgianni et al., 2020; Ogawa et al., 2019). Another example is the 3FTx in Elapidae, which may have evolved from the gene of GPI-anchored Ly6 protein by losing the sequence encoding the peptide responsible for the attachment to GPI lipids on the membrane (Tsetlin, 2015).

Aside from the VSTs mentioned above, the low quantity of minor venom components hinders researchers from studying them, rendering their biological significance elusive. One minor venom component, snake venom phosphodiesterase (svPDE), has been found to be ubiquitous in the venoms of most venomous species in Viperidae (~0.01 – 2.5%) (Damm, Hempel, & Sussmuth, 2021) and Elapidae (~0.4 – 1.1%) (Laustsen et al., 2015; C. H. Tan, Tan, Fung, & Tan, 2015). SvPDE is a soluble high-molecular-weight glycoprotein and is distinct from the intracellular phosphodiesterase (Al-Saleh & Khan, 2011; Mitra & Bhattacharyya, 2014; Oliveira et al., 2021; L. Peng et al., 2011; Santoro, Vaquero, Paes Leme, & Serrano, 2009; Trummal et al., 2014; Valerio, Corradini, Panunto, Mello, & Hyslop, 2002). Since extracellular ATP is involved in epithelial homeostasis (Mori et al., 2022) and also functions as a danger signal of damaged cells through the purinergic

signaling pathway (Burnstock, 2016; Cintra-Francischinelli et al., 2010), svPDE, which enzymatically acts on the extracellular ATP, is expected to perturb the related physiological responses. However, the gene encoding svPDE, the evolutionary origin of that gene and the potential biological targets of svPDE remain largely unknown. In this study, we integrated comparative genomics, comparative transcriptomics of the venom glands, and the biochemical analysis of svPDE obtained directly from the venom, to investigate the genomic locus, gene structure, evolution, protein structure, enzymatic activities and potential targets of svPDE from *Naja atra*.

Results

The evolution of the soluble secretory svPDE from the membrane-anchored ENPP3 via alternative splicing of their encoding genes

The presence of svPDE in venoms has been identified by proteomics or by detecting its enzymatic activity for decades. However, its gene characteristics have not been well reported. To reveal the genomic locus and the gene structure of svPDE, we integrated genomic, transcriptomic and proteomic data. The query peptide sequence of svPDE was retrieved from PDB database (accession: 5GZ4). This svPDE was directly purified from the crude venom of *Naja atra* captured in Taiwan. The target genomes included the draft one of *Naja atra* from our ongoing internal project and the complete one of its sister species, *Naja naja* (Suryamohan et al., 2020). For genomes of *Naja atra* and *Naja naja*, the only genomic hit of the svPDE peptide sequence is located on the ENPP3 locus inferred from the conserved synteny, while the peptide sequence is a subsequence of the translated sequence inferred from the putative exons of the ENPP3 gene **(*SI Appendix*, Fig. S1)**. Consistently, compared with the ENPP3 peptide sequence, the purified svPDE is shorter at the N-terminal, lacking the cytoplasmic (CP) and transmembrane (TM) domains of ENPP3 **(*SI Appendix*, Fig. S1)**. The unique genomic hit, conserved synteny, and identical amino acid sequences strongly suggest that svPDE and ENPP3 are encoded from the same genomic locus and share all but the 5' end exons.

The absence of CP and TM domains in purified svPDE protein is consistent with the secretory feature of most venom proteins **(Fig. 1A** and ***SI Appendix*, Fig. S1)**. However, as the ENPP3 gene is highly conserved in the entire Metazoan clades **(*SI Appendix*, Fig. S2)** and is essential for the basic biochemical functions (Borza, Salgado-Polo, Moolenaar, & Perrakis, 2021), instead of replacing original ENPP3 by svPDE, their transcripts are expected to coexist in snakes. One possibility for their coexistence is through an alternative splicing mechanism (Sorek, 2007; Verta & Jacobs, 2022), in which alternative 5' exons with a signal peptide emerged, leading to the soluble secretory form without damaging the functions of the remaining peptides. To test our hypothesis, we comprehensively *de novo* assembled transcriptomes from the species across 13 clades of Toxicofera **(Fig. 1B)** with publicly available RNA-Seq data and compared them with the

corresponding genomes available in the NCBI Assembly database **(*SI Appendix*, Table S1)**. In addition to the transcript encoding the conserved ENPP3 protein, we found potential svPDE transcripts comprising all but the first (E1) exons of the ENPP3 gene.

A careful comparison reveals two types of novel clade-specific E1 of svPDE transcripts **(Fig. 1B** and ***SI Appendix*, Fig. S3)**. The first one, named svPDE-E1a, found in the transcriptomes of all snake clades but Elapidae and Hydrophiidae; and the other, named svPDE-E1b, found only in Elapidae and Hydrophiidae. Both svPDE-E1a and svPDE-E1b accompany a putative TATA-box core promoter located on their proximal upstream **(Fig. 1A** and ***SI Appendix*, Fig. S4)** and encode signal peptides necessary for the secretory feature of svPDE **(Fig. 1C** and ***SI Appendix*, Fig. S3)**. In addition, the canonical GT-AG introns between the novel E1 (for both two types) and the conserved E2 were consistently observed in all snake species with available genomes. Together, our results suggest that the svPDE protein came from the co-option of the ancestral ENPP3 gene by using a novel 5' exon.

The recruitment of svPDE

We next traced the origin of the novel E1 in Toxicofera by examining transcriptomic and genomic data. We found that the genomic sequences of svPDE-E1a were present and conserved in all species of Serpentes (including Elapidae and Hydrophiidae, in which the expressed transcripts do not consist of svPDE-E1a) but not in the species of Dactyloidae, Agamidae, Varanidae, and Typhlopidae. This suggests an early emergence of svPDE-E1a in the common ancestor of Serpentes and became non-functional in some descent lineages (i.e., not expressed in Pythonidae, Elapidae and Hydrophiidae) **(Fig. 1B)**. Pythonidae has genomic mutations in the TATA-box of svPDE-E1a **(Fig. 1C** and ***SI Appendix*, Fig. S4A)** and the absence of an alternative mechanism to express svPDE is probably associated with its non-venomous character (Reyes-Velasco et al., 2015). For svPDE-E1a in some species of Dipsadidae and Lamprophiidae, although their original translation start codons (ATG) had mutated, alternative ATGs were recruited at the downstream region of the original ones **(Fig. 1C** and ***SI Appendix*, Fig. S4B)**. This recruitment of start codons was caused by a newly 3' splicing site of svPDE-E1a emerged in their common ancestor **(*SI Appendix*, Fig. S5)** that elongates the length of svPDE-E1a at the 3' end. This replacement allows for the possibility of svPDE translation in Dipsadidae and Lamprophiidae.

Similar replacements of elements involved in svPDE expression were also found in Elapidae and Hydrophiidae: the emergence of exon svPDE-E1b. Although svPDE-E1a is highly conserved in the genomes of Elapidae and Hydrophiidae, it was not found in their transcriptomes, potentially due to the genomic mutation of a TATA-box at the putative core promoter for svPDE-E1a in their common ancestor **(Fig. 1C** and ***SI Appendix*, Fig. S4A)**. Instead, their svPDE transcripts include another 5' exon, svPDE-E1b, which was not

identified in the genomes of other clades that diverged earlier. These results indicate that svPDE transcripts switched from using svPDE-E1a to svPDE-E1b in the common ancestor of Elapidae and Hydrophiidae **(Fig. 1C)**.

Notably, the entire svPDE-E1b, including its putative promoter, is embedded in the antisense strand of an uncharacterized Elapidae mobile element (Dfam accession: DR0148352, ***SI Appendix*, Fig. S4C**) that inserted into the ENPP3 locus in the common ancestor of Elapidae and Hydrophiidae **(Fig. 1C)**. We examined other DR0148352 elements in the genome-available species of Elapidae and Hydrophiidae and found that the homologous promoter region in the elements do not contain TATA-box patterns or other promoter-related domains **(*SI Appendix*, Fig. S4C**). This suggests that the putative promoter of svPDE-E1b emerged *in situ* after the mobile element insertion. Interestingly, the efficiency of svPDE expression could be associated with the TATA-box of svPDE-1b since Elapidae and Hydrophiidae express significantly high and low levels of svPDE **(Fig. 2A)** in line with the canonical and non-canonical TATA-box patterns **(*SI Appendix*, Fig. S4C)**, respectively.

SvPDE are predominantly expressed in the venom glands of venomous snakes

We then compared the svPDE expression in venom glands among different snake clades using transcriptomic data. As the sequences of svPDE and ENPP3 only differ in the first exon, the reads mapped to the shared exons cannot be distinguished from different transcripts. Thus, we estimated the relative expression of svPDE over ENPP3 by the difference in the number of reads spanning distinct E1-E2 junctions of ENPP3 and svPDE splicing forms. This method also reduced the potential biases caused by different experiments, independent samples, the uncertainty of 5' UTR length due to library preparation, and the uneven distribution of mapped reads.

As shown in **Fig. 2A**, for the relative differences of svPDE over ENPP3 expression, most species of Viperidae, Lamprophiidae and Elapidae exhibit a predominance of svPDE in their venom glands, in line with the current understanding of venoms by proteomic studies in such clades (Damm, Hempel, Nalbantsoy, & Sussmuth, 2018; Gren et al., 2019). The relative expression of svPDE was inferior to ENPP3 in the glands of Homalopsidae, Colubridae, Dipsadidae and Hydrophiidae **(Fig. 2A)**. Intriguingly, compared to the deficient expression in the other three clades, the expression of svPDE in Homalopsidae was higher, although not larger than ENPP3. This feature provides new insight into the characteristics of Homalopsidae, a species still being discovered (Bernstein, Murphy, Voris, Brown, & Ruane, 2021; Kohler et al., 2021). In contrast to the predominance of svPDE across the species of Viperidae, the svPDE expression across the species of Elapidae shows higher interspecies variation and even exhibits a contrary pattern in some of the species **(Fig. 2A)**. Proteomic studies have identified this variation in Elapidae for decades (Modahl, Roointan, Rogers, Currier, & Mackessy, 2020; N. H. Tan &

Tan, 1988), but it still requires future functional studies of svPDE in the venoms to understand the significance of this variation.

The tissue-specific expression of svPDE

Traditionally, svPDE studies have only focused on the venom gland by detecting its enzyme activity and neglected the expression in other tissues. We identified the difference between svPDE and ENPP3 transcripts. The reads originated from two transcripts which became distinguishable in the transcriptomic data. We investigated the relative expression of svPDE over ENPP3 in other tissues to understand their tissue-specificity and whether they are mutually exclusive. In both Elapidae **(Fig. 2B),** and Viperidae **(Fig. 2C-D)**, taking India cobra (*Naja naja*), habu (*Protobothrops flavoviridis*) and prairie rattlesnake (*Crotalus viridis*) as examples, venom glands exhibit excessively high svPDE levels and very low ENPP3 expression. On the other hand, ENPP3 is primarily abundant in small intestines, and for those tissues enriched with ENPP3, the frequency of svPDE transcripts is low or nonexistent. These results indicate the mutually exclusive relationship between svPDE and ENPP3. Interestingly, besides the venom gland, prairie rattlesnakes show a predominant expression of svPDE in the kidney **(Fig. 2D)**.

The structure of svPDE and its binding targets

Even though the presence of phosphodiesterase enzymatic activities against ATP, ADP and/or DNA have been demonstrated in many crude venoms, the exact structural element responsible for such activities is still unclear. We therefore purified the svPDE directly from the crude venom of Taiwan cobra (*Naja atra*) via chromatography **(*SI Appendix*, Fig. S6)** and determined its 3D structure with X-ray crystallography. The structures of unliganded svPDE (apo form, PDB accession: 5GZ4) and liganded svPDE (AMP-complexed form, PDB accession: 5GZ5) were determined at resolutions of 2.55 and 2.09 Å **(*SI Appendix*, Table S2)**, respectively. Crystal structures show that svPDE resembles human ENPP3 and ENPP1 proteins (the structural homologs with currently available structural data) with *N*-terminal somatomedin domains (SMB1 and SMB2), a catalytic phosphodiesterase domain (PDE) and a C-terminal nuclease domain (NUC) **(Fig. 3A** and **Fig. 3B)**. An insertion loop (IL) and two catalytical zinc ions constitute the bimetallic active site of the catalytic PDE domain **(Fig. 3C)**. AMP is accommodated in a nucleotide-binding pocket formed by residues W251/E255/Y269 from the IL loop and residues F186/K271 from the catalytic PDE domain **(Fig. 3C** and **Fig. 3D)**, where T185 is the nucleophile residue responsible for the enzymatic activity. The conformation of the active site in the apo form is almost identical to that in the AMP-complexed form. When AMP is bound to svPDE, only side chains of N206 and K271 shift and interact with the O atom of the phosphate group and the N6 atom of AMP, respectively **(Fig. 3C)**.

Biochemical measurements of the substrate specificity revealed that svPDE could hydrolyze ATP, ADP and NAD (nicotinamide adenine dinucleotide) with relatively low K_m

compared to the hydrolysis of GTP, CTP and UTP **(*SI Appendix*, Table S3)**, indicating that svPDE is a more effective catalyst for adenine-containing nucleotides. Given that the 3D structure and substrates are conserved for both svPDE and ENPP3, svPDE could hydrolyze extracellular nucleotides, showing the preference for adenine nucleotides and derivatives. Partial electron densities of N-glycans show that seven residues of svPDE have been post-translationally modified **(Fig. 3E and *SI Appendix*, Fig. S7)**. N512, one of the seven N-glycans, and C408-C795 disulfide linkage are essential for the stabilization of the PDE-NUC interface of ENPP members and were found to be conserved in svPDE.

Apart from the enzymatic activities, membrane-attached human ENPP1 has been shown to be able to inhibit insulin receptor (IR) signaling processes (Kato et al., 2012), and membrane-attached human ENPP3 plays a functional role in the Golgi apparatus of neuronal cells to suppress the activity of 1,6-N-acetylglucosaminyltransferase, GnT-IX (Korekane et al., 2013) and perturbed N-glycosylation functions. Interestingly, mass spectrometric analysis of the N-glycosylation patterns with less terminal sialic acid contrasts sharply with our previous study on venom glycoproteins such as SVMP (Huang et al., 2015). Without terminal sialic acids, the exposed lactose disaccharide domain of the N-glycan moiety in svPDE enables its binding to immunologically important galectins. As shown in **Fig. 3F** and **Fig. 3G**, svPDE indeed interacts with IR ectodomain and Gal-3 with apparent binding affinities K_D~1.8 μM and ~269 μM, respectively. A recent study reported that direct binding of Gal-3 to insulin receptors triggers the disease state of mice adipocytes (P. Li et al., 2016). Hence, it would be interesting to investigate how svPDE might interact with IR and Gal-3 and get involved in the tissue damage and enhancement of cell toxicity resulting from VSTs, such as 3FTx and cytotoxins, in cobra snakebites.

Discussion

Since ENPP3 is shared among Metazoan species, but svPDE is limited to snakes **(*SI Appendix*, Fig. S2)**, our result suggests that svPDE has evolved from an ancestral ENPP3 gene by co-option, an evolutionary strategy of using pre-existing proteins for new functions. This type of co-option was also used for other minor venom components of 5' nucleotidase (5NT), as shown in a recent study on the *Bothrops jararaca* genome (Almeida et al., 2021). By using an alternative 5' exon, ENPP3 replaced its domains for membrane attachment by the signal peptide and became a secretory svPDE protein. The mechanism of svPDE expression has changed multiple times during snake evolution, and the one that emerged in the common ancestor of Elapidae and Hydrophiidae was associated with an insertion of mobile elements **(Fig. 1C)**. New genes evolving from the insertion of mobile elements have been found in various species, such as the genes that arose through L1 and Alu insertion in primates (Long, 2001). Based on the transcriptomic data, this mobile element is still active and the genomic changes caused by its insertion may still be happening in the extant species. Such expansion of mobile elements has been

observed and proposed to contribute to the adaptation of Hydrophiidae recently (C. Peng et al., 2020).

Notably, the replacement of regulatory elements that evolved multiple times during snake evolution strongly suggests the importance of keeping or improving the expression of svPDE in the mechanism of venom action. Nevertheless, different replacements of exons and regulatory elements that emerged in Viperidae and Elapidae clades could support and be probably associated with the idea that the fangs of these two clades have evolved independently (Westeen, Durso, Grundler, Rabosky, & Davis Rabosky, 2020).

It should be noted that seven ENPP family members have merged and duplicated into eight copies (two ENPP7 copies) in the ancestor of Gnathostomata, i.e., before the divergence of Acanthodians (cartilaginous fishes) and Euteleostomi (the common ancestor of other bony vertebrates), and then followed by a loss of ENPP5 in the snake lineage **(*SI Appendix*, Fig. S2)**. Since snake genomes contain seven ENPP copies (except for the loss of ENPP5) sharing the conserved PDE domain, all copies may be reported as the targets while using svPDE peptide as a query against the genomes and thus may cause the overestimation of svPDE gene copies in other studies (Rao et al., 2022).

Comparative transcriptomics is a powerful tool to reveal species-specific or tissue-specific novel transcripts. For example, the expression of svPDE in the less-known Lamprophiidae, which has not been reported, is probably associated with the adaptation of prey selection as a dietary generalist compared to Viperidae and Elapidae (Portillo et al., 2019). In addition, although the svPDE transcripts are rare in the glands of Colubridae, Dipsadidae, and Hydrophiidae, the junction spanning reads and/or *de novo* assembled transcripts indicate their existence in such clades. Although low, the svPDE expression in Duvernoy's glands of Colubridae is intriguing as Colubridae is traditionally regarded as a non-venomous clade. Furthermore, as Duvernoy's glands seem to be a replacement for the venom gland from the perspectives of evolution and anatomy, its svPDE expression might imply some unknown function of Duvernoy's glands in Colubridae. Similar questions are also worthy to be addressed for the rictal glands of pythons, which are believed to be a relic of the secretion system used by ancestral snakes. Furthermore, the predominance of svPDE in prairie rattlesnakes' kidneys **(Fig. 2D)** was only found in males, suggesting an unknown physiological role of svPDE in the kidneys of male vipers. It would be interesting to further explore this sexual bias when more data becomes available from different species and individuals.

Based on our biochemical analysis, similar to ENPP3, svPDE can hydrolyze a variety of substrates, including nucleotides and nucleotide derivatives that can induce multiple cellular effects on the venom-exposed tissues. ENPP3 has been reported to downregulate extracellular ATP secreted from basophils and mast cells and to suppress allergic

inflammation (Tsai et al., 2015), raising a question of what function has adopted from membrane-attached ENPP3 to secretory svPDE. In the cases of venomous snakebites, extracellular ATP released by damaged cells can activate complex physiological responses, such as platelet aggregation, mast cell secretion, inflammation, membrane permeability, vascular function and neurotransmission (Gordon, 1986), and can associate with purinergic receptors as a danger signal to initiate immune responses. Another nucleotide metabolite, NAD, is also believed to activate purinergic receptors to alert the immune system if released to the extracellular matrix (Audrito, Messana, Brandimarte, & Deaglio, 2021; Haag et al., 2007). Together, it is conceivable that the removal of extracellular ATP and NAD by svPDE alters the purinergic signaling of cells around the bitten tissues and redirects the immune defense mechanism of the prey in responding to venom actions.

Despite the catalytic functions of svPDE, we also demonstrated that svPDE is capable of binding to the IR and Gal-3, suggesting a novel physiological role for svPDE. Previous studies have revealed that ENPP1 directly interacts with IR and inhibits insulin signaling (Kato et al., 2012; Maddux & Goldfine, 2000), and the binding of Gal-3 to IR causes insulin resistance (P. P. Li et al., 2016). In similar fashion, the binding of svPDE to the insulin receptor and Gal-3 may result in the perturbation of insulin signaling. Although snakebite symptoms ascribed to the insulin insensitivity and impaired glucose metabolism need to be further investigated, the versatile role of svPDE in cellular metabolism and immune modulation draws attention to the different aspects of venom action.

The co-opted secretory svPDE in snake venom likely evolved to reuse the original intracellular functions of ENPP3 at the extracellular matrix of snakebite wounds. We speculated that, around the damaged tissues, svPDE interferes with the normal cellular physiology in which ENPP family are involved. Although the exact functions and physiological significance of svPDE are still unclear, it reveals that even as a minor venom component, it may play a much more complicated role in toxicity by perturbing signal transduction that was neglected before. In the light of the ubiquitous presence of svPDE in most, if not all, snake venoms, the highly conserved svPDE is a candidate target in developing a generic antidote for venomous snakebites across different species clades. In contrast to the current emergency room treatment for snakebites that relies only on anti-snake venom serums, the generic drug would significantly reduce the burden of snakebites on human health.

Materials and Methods

Whole-genome assembly of *Naja atra* habitated in Taiwan

DNA was extracted from the muscle tissue of *Naja atra* by using AllPure Genomic DNA Extraction Kit (AllBio Inc, Taiwan). Third-generation sequencing was obtained by using the

circular consensus sequencing (CCS) mode on the PacBio Sequel II System resulting in 4563622 HiFi pass reads (56.8 Gb with an average length of 12.4 Kb). Next-generation sequencing was obtained in PE150 on the Illumina NovaSeq 6000 System and pre-processed by fastp (version 0.19.4) (S. Chen, Zhou, Chen, & Gu, 2018) to discard reads with ambiguous N and trim five nucleotides on the both ends of read, resulting in 559566268 clean paired-reads (~78Gb). PacBio HiFi filtered reads were used for genome assembly by hifiasm (version 0.13-r308) (Cheng, Concepcion, Feng, Zhang, & Li, 2021), resulting in a draft genome in a size of 1.88Gb and N50 of 29.7Mb). Assembled contigs were then polished by Nextpolish (version 1.3.1) (Hu, Fan, Sun, & Liu, 2020) with Illumina clean paired-reads. See ***SI Appendix*, Tables S4, S5 and S6** for detailed statistics.

Revealing the ENPP3 genomic exons

The ENPP3 exons were identified in the species with available genomes by an iterative search strategy (Pan & Lin, 2020) with HMMER (version 3.2.1) (Wheeler & Eddy, 2013). Once the hit was found in a species, its sequence was used to update the HMM profile; as the sequences from more species were involved in a profile, it could more sensitively identify the homologs in distant species. The identified exons were scrutinized for the splicing junctions and cross-checked by aligning with *de novo* assembled transcripts.

Identification of svPDE and ENPP3 transcripts

Publicly available RNA-Seq reads were processed for quality and adapter trimming by using fastp (version 0.21.0) (S. Chen et al., 2018) and then *de novo* assembled to reconstruct the transcriptome using Trinity (version 2.13.2) (Grabherr et al., 2011). The potential svPDE transcript was selected by removing the contigs that well-aligned to the concatenated joined sequence of conserved ENPP3 exons. Fragmented contigs aligned as substrings were excluded by the consideration that they are assembled with sequencing error reads resulting in failure of extension. Since the precise position of the transcription start site is undeterminable due to the nature of poly-A based library preparation, the longest transcript at the 5' end was used for positioning the genomic upstream region of promoter prediction.

Prediction of promoters and mobile elements

Transmembrane domain and signal peptide were predicted by TMHMM (version 2.0) (Y. Chen, Yu, Luo, & Jiang, 2003) and SignalP (version 5.0) (Almagro Armenteros et al., 2019), respectively. The genomic upstream of the transcripts was fetched for searching promoters. The core promoter was predicted by ElemeNT (Sloutskin et al., 2015) (version 2). The mobile element DR0148352 was recognized by the exceeding multiple hits of such region in the genomes during the iterative search strategy for exons and annotated by searching the Dfam database (release 3.5) (Storer, Hubley, Rosen, Wheeler, & Smit, 2021).

Estimation of svPDE and ENPP3 expressions

The counting of E1-E2 junction spanning reads was conducted instead of the typical mapping approach because ENPP3 and svPDE transcripts share long identical sequences (24 exons), from which the original transcripts of reads cannot be distinguished. Meanwhile, the imprecise end of 5' UTR during sequencing might influence the unfair mapping of the first exon across samples. Thus, the expression level of svPDE and ENPP3 transcripts was estimated by directly counting the reads spanning at least 30 bp (i.e., 15 bp in each exon) at the exon-exon junctions between specific E1 (ENPP3-E1, svPDE-E1a and svPDE-E1b) and shared E2. In order to avoid the estimation biases caused by different studies, experiments and individuals, the estimation was calculated by the log2 ratio of read counts for svPDE over ENPP3 transcripts within a sample (i.e., a fold change of svPDE over ENPP3 expression). A count of zero was adjusted to a pseudo-count of one to prevent the undefinable log transformation of zero or infinitive values.

Purification of svPDE from the crude venom of *Naja atra*

One gram of crude venom was dissolved in 3 mL of 50 mM phosphate buffer (pH 6.2) with 500 mM NaCl, filtered and applied to a Sephadex G-75 column (General Electric Company), pre-equilibrated with 50 mM phosphate buffer (pH 6.2). The fraction containing svPDE (the peak highlighted in red in ***SI Appendix*, Fig. S1A**) was collected and applied to a MonoQ column (General Electric Company). The flow-through fraction (the peak highlighted in red in ***SI Appendix*, Fig. S1B**) was concentrated, dialyzed against 50 mM phosphate buffer (pH 7.4) and applied to a Hitrap Heparin column (General Electric Company). The bound proteins were eluted with 50 mM phosphate buffer (pH 7.4) with a gradient of NaCl rising from 0 to 1M. The first eluted fraction (the peak highlighted in red ***SI Appendix*, Fig. S1C**) was concentrated, dialyzed against 35 mM phosphate buffer (pH 7.2), and applied to a MonoS column (General Electric Company). Subsequently, 35 mM phosphate buffer (pH 7.2) with a gradient of NaCl from 0 to 1 M was used to elute the bound proteins. The fraction following the flow-through fraction (the peak highlighted in red in ***SI Appendix*, Fig. S1D**) was concentrated and subjected to a Superdex 200 column (General Electric Company), pre-equilibrated in 20 mM HEPES (pH 7.4) with 100 mM NaCl. Molecular weight and the purity of svPDE were analyzed with a reducing 10% (w/v) SDS/PAGE followed by staining with Coomassie blue (inset in ***SI Appendix*, Fig. S5E**).

Protein crystallization

Crystallization of svPDE was performed at 293K by the hanging drop vapor-diffusion method. Pure protein solution (12 mg/ml) was mixed with a well solution containing 0.1 M imidazole (pH 6.6), 0.2 M zinc acetate and 23 % PEG 3350 in 1:1 volume ratio. Single crystals were grown to ~0.1 mm in the longest dimension after 15 days. To obtain crystals of svPDE in complex with AMP, the svPDE crystal, grown by mixing 1 μl of protein solution (12mg/ml) with 1 μl of well solution (0.1 M imidazole pH 6.5, 0.2 M zinc acetate and 20 %

PEG 3000), was soaked with 1 mM AMP for 2 days. SvPDE crystals of apo form and AMP-complexed form were cryoprotected by a brief transfer to their respective reservoir solutions supplemented with 20% ethylene glycol before data collection.

Structure determination and refinement

The data sets of svPDE in the apo form (unliganded) and the AMP-complexed form were collected at BL13C1 and BL15A1 of NSRRC (National Synchrotron Radiation Research Center, Taiwan), respectively. Both data sets were indexed, integrated, and scaled with HKL2000. The structure was solved by molecular replacement with the program Phaser (McCoy et al., 2007) using the PDE and NUC domains from mouse ENPP1 (PDB accession: 4B56) as the search model. SMB1 and SMB2 domains from mouse ENPP2 (PDB accession: 3NKM) were used as templates to build the N-terminal structure. Subsequently, structural refinements were performed using Coot (Emsley, Lohkamp, Scott, & Cowtan, 2010) and PHENIX (Liebschner et al., 2019). Crystallographic and refinement statistics are listed in ***SI Appendix*, Table S2**. All structure figures were obtained using PyMOL (Version 1.3, Schrödinger, LLC) and Chimera (Pettersen et al., 2004). The structure in the AMP-complexed form was solved by using the apo form of svPDE as a search model. However, there is no clear electron density map allowing us to build the SMB1 domain of svPDE in the AMP-complexed form, indicating the SMB1 domain could be highly flexible. B factor representations of svPDE in apo and AMP-complexed forms identified the N-terminus including SMB1 and SMB2 as a highly flexible region **(*SI Appendix*, Fig. S9)**.

Nucleotide and NAD degradation assay

The nucleotide and NAD degradation activities of svPDE from *Naja atra* were analyzed by an HPLC equipped with an analytical C18 column. The reaction mixture containing 40 nM (for nucleotide degradation assay) or 80 nM (for NAD degradation assay) svPDE and 1 mM ATP or ADP or NAD in 20 mM Tris-HCl (pH 8.0) was incubated at 37 °C. After various time periods, reactions were terminated by adding 1.0 N NaOH and analyzed by HPLC. In addition, enzyme kinetics of svPDE for nucleotides and NAD were also determined using HPLC. 40 nM (for nucleotide degradation) or 80 nM (for NAD degradation) svPDE was incubated with various concentrations of ATP or ADP or NAD in 20 mM Tris-HCl (pH 8.0) at 37 °C. The reactions were terminated by addition of 1.0 N NaOH and analyzed by HPLC. K_m and V_{max} were estimated from a Michaelis Menten plot of the initial rates as a function of substrate concentrations. *k*cat was obtained using the equation: *k*cat = V_{max} / *[E].*

Mass spectrometric analysis

Trypsin digestion of svPDE from *Naja atra* was performed at an enzyme to substrate ratio of 1:50 (w/w) after the cysteine residues of svPDE were reduced and alkylated with dithiothreitol and iodoacetamide, respectively. Glycopeptides were enriched by hydrophilic interaction (HILIC) liquid chromatography using Superdex 75 prep grade resins. The

binding and wash buffers for HILIC were 75% acetonitrile (ACN), 0.1% formic acid (FA). Glycopeptides were eluted from the resins with 50% and then 25% ACN in the presence of 0.1% FA. The dried glycopeptide mixtures were reconstituted with H_2O, 0.1% FA and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). The glycopeptides were separated on a reverse phase column (BEH C18, 0.1 x 100 mm, Waters Cooperation) at a flow rate of 300 nL/min using a 60-min ACN gradient in the presence of 0.1% FA. The effluents were analyzed online by an Orbitrap Elite hybrid mass spectrometer (Thermo Fisher Scientific Inc.). The mass spectrometer was operated in positive ion mode, and the tandem mass spectra (MS/MS) were acquired in a data-dependent manner. Briefly, the six most intense ions in the survey MS spectrum were selected for collision induced dissociation (CID) and then scanned in the orbitrap analyzer for obtaining the MS/MS spectra. The collected MS/MS spectra were subjected to glycopeptide identification using the Byonic software (Protein Matric, inc.). The database search was performed against a focused database containing four venom glycoproteins (svPDE, L-amino acid oxidase, 5'-nucleotidase and cobra venom factor) with the following parameters: enzyme specificity, fully tryptic; maximum number of missed cleavages, 2; fixed modifications, carbamidomethylation of cysteine; variable modifications, oxidation of methionine, acetylation of protein N-terminus, and glutamine to pyroglutamate conversion at peptide N-terminus; N-glycans, 309 mammalian N-glycans provided by the Byonic software; mass tolerances, 15 ppm and 100 ppm for MS and MS/MS spectra, respectively. Finally, the matched proteins were filtered to 1% false discovery rates (FDRs) estimated by a decoy database search.

Solid-phase binding assay

The flat-bottomed, 394-well microliter plate was coated with 50 μg/ml svPDE in the presence of coating buffer (0.1M NaHCO3, pH 9.4). After 3 hours of incubation at 37 °C, the free surface of the wells was blocked by adding 5% BSA/PBS and the plate was incubated at 37 °C for an additional 2 hours. Human recombinant Galectin 3 (His tag) (Sino Biological Inc.) in PBS was then applied to the wells coated with svPDE after the wells were washed with PBST three times. The soluble Gal3-His was then allowed to interact with immobilized svPDE at 37 °C for 16 hours. In order to remove the unbound Gal3-His after reaction, the wells underwent three PBST washes. A mouse 6X His tag antibody (HRP) (GeneTex Inc.) was then used to recognize Gal3-His at room temperature for 1 hour. After washing with PBST for 3 times to eliminate nonspecific interactions, bound Gal3-His was quantified by measuring the absorbance at 450 nm after adding the substrate, 3,3',5,5'-Tetramethylbenzidine (TMB) for horseradish and 1N H2SO4 for termination of the reaction.

Acknowledgments

This research was funded by Higher Education SPROUT Project launched by the Ministry of Education, Taiwan (R.O.C.). CTP and HHC were supported by Yushan Scholar

Program. We thank the experimental facility and the technical services provided by the "Synchrotron Radiation Protein Crystallography Facility of the National Core Facility Program for Biotechnology, Ministry of Science and Technology" and the "National Synchrotron Radiation Research Center", a national user facility supported by the Ministry of Science and Technology, Taiwan (R.O.C.). We are also immensely grateful to Gerry Tonkin-Hill and Jessica K. Calland for improving the manuscript.

References

Al-Saleh, S. S., & Khan, S. (2011). Purification and characterization of phosphodiesterase i from Walterinnesia aegyptia venom. *Prep Biochem Biotechnol, 41*(3), 262-277. doi:10.1080/10826068.2011.575319

Almagro Armenteros, J. J., Tsirigos, K. D., Sonderby, C. K., Petersen, T. N., Winther, O., Brunak, S., . . . Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol, 37*(4), 420-423. doi:10.1038/s41587-019-0036-z

Almeida, D. D., Viala, V. L., Nachtigall, P. G., Broe, M., Gibbs, H. L., Serrano, S. M. T., . . . Junqueira-de-Azevedo, I. L. M. (2021). Tracking the recruitment and evolution of snake toxins using the evolutionary context provided by the Bothrops jararaca genome. *Proc Natl Acad Sci U S A, 118*(20). doi:10.1073/pnas.2015159118

Audrito, V., Messana, V. G., Brandimarte, L., & Deaglio, S. (2021). The Extracellular NADome Modulates Immune Responses. *Frontiers in Immunology, 12*. doi:ARTN 704779 10.3389/fimmu.2021.704779

Bernstein, J. M., Murphy, J. C., Voris, H. K., Brown, R. M., & Ruane, S. (2021). Phylogenetics of mud snakes (Squamata: Serpentes: Homalopsidae): A paradox of both undescribed diversity and taxonomic inflation. *Mol Phylogenet Evol, 160*, 107109. doi:10.1016/j.ympev.2021.107109

Borza, R., Salgado-Polo, F., Moolenaar, W. H., & Perrakis, A. (2021). Structure and function of the ecto-nucleotide pyrophosphatase/phosphodiesterase (ENPP) family: Tidying up diversity. *J Biol Chem, 298*(2), 101526. doi:10.1016/j.jbc.2021.101526

Burnstock, G. (2016). Purinergic Mechanisms and Pain. *Adv Pharmacol, 75*, 91-137. doi:10.1016/bs.apha.2015.09.001

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics, 34*(17), i884-i890. doi:10.1093/bioinformatics/bty560

Chen, Y., Yu, P., Luo, J., & Jiang, Y. (2003). Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT. *Mamm Genome, 14*(12), 859-865. doi:10.1007/s00335-003-2296-6

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods, 18*(2), 170-175. doi:10.1038/s41592-020-01056-5

Cintra-Francischinelli, M., Caccin, P., Chiavegato, A., Pizzo, P., Carmignoto, G., Angulo, Y., . . . Montecucco, C. (2010). Bothrops snake myotoxins induce a large efflux of ATP and potassium with spreading of cell damage and pain. *Proc Natl Acad Sci U S A, 107*(32),

14140-14145. doi:10.1073/pnas.1009128107

Damm, M., Hempel, B. F., Nalbantsoy, A., & Sussmuth, R. D. (2018). Comprehensive Snake Venomics of the Okinawa Habu Pit Viper, Protobothrops flavoviridis, by Complementary Mass Spectrometry-Guided Approaches. *Molecules, 23*(8). doi:10.3390/molecules23081893

Damm, M., Hempel, B. F., & Sussmuth, R. D. (2021). Old World Vipers-A Review about Snake Venom Proteomics of Viperinae and Their Variations. *Toxins (Basel), 13*(6). doi:10.3390/toxins13060427

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 4), 486-501. doi:10.1107/S0907444910007493

Giorgianni, M. W., Dowell, N. L., Griffin, S., Kassner, V. A., Selegue, J. E., & Carroll, S. B. (2020). The origin and diversification of a novel protein family in venomous snakes. *Proc Natl Acad Sci U S A, 117*(20), 10911-10920. doi:10.1073/pnas.1920011117

Gordon, J. L. (1986). Extracellular Atp - Effects, Sources and Fate. *Biochemical Journal, 233*(2), 309-319. doi:DOI 10.1042/bj2330309

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., . . . Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol, 29*(7), 644-652. doi:10.1038/nbt.1883

Gren, E. C. K., Kitano, E. S., Andrade-Silva, D., Iwai, L. K., Reis, M. S., Menezes, M. C., & Serrano, S. M. T. (2019). Comparative analysis of the high molecular mass subproteomes of eight Bothrops snake venoms. *Comp Biochem Physiol Part D Genomics Proteomics, 30*, 113-121. doi:10.1016/j.cbd.2019.01.012

Gutierrez, J. M., Calvete, J. J., Habib, A. G., Harrison, R. A., Williams, D. J., & Warrell, D. A. (2017). Snakebite envenoming. *Nat Rev Dis Primers, 3*, 17063. doi:10.1038/nrdp.2017.63

Haag, F., Adriouch, S., Brass, A., Jung, C., Moller, S., Scheuplein, F., . . . Koch-Nolte, F. (2007). Extracellular NAD and ATP: Partners in immune cell modulation. *Purinergic Signalling, 3*(1-2), 71-81. doi:10.1007/s11302-006-9038-7

Hu, J., Fan, J., Sun, Z., & Liu, S. (2020). NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics, 36*(7), 2253-2255. doi:10.1093/bioinformatics/btz891

Huang, H. W., Liu, B. S., Chien, K. Y., Chiang, L. C., Huang, S. Y., Sung, W. C., & Wu, W. G. (2015). Cobra venom proteome and glycome determined from individual snakes of Naja atra reveal medically important dynamic range and systematic geographic variation. *J Proteomics, 128*, 92-104. doi:10.1016/j.jprot.2015.07.015

Kato, K., Nishimasu, H., Okudaira, S., Mihara, E., Ishitani, R., Takagi, J., . . . Nureki, O. (2012). Crystal structure of Enpp1, an extracellular glycoprotein involved in bone mineralization and insulin signaling. *Proc Natl Acad Sci U S A, 109*(42), 16876-16881. doi:10.1073/pnas.1208017109

Kohler, G., Khaing, K. P. P., Than, N. L., Baranski, D., Schell, T., Greve, C., . . . Pauls, S. U. (2021). A new genus and species of mud snake from Myanmar (Reptilia, Squamata, Homalopsidae). *Zootaxa, 4915*(3), zootaxa 4915 4913 4911. doi:10.11646/zootaxa.4915.3.1

Korekane, H., Park, J. Y., Matsumoto, A., Nakajima, K., Takamatsu, S., Ohtsubo, K., . . . Taniguchi, N.

606 (2013). Identification of ectonucleotide pyrophosphatase/phosphodiesterase 3 (ENPP3) as
607 a regulator of N-acetylglucosaminyltransferase GnT-IX (GnT-Vb). *J Biol Chem, 288*(39),
608 27912-27926. doi:10.1074/jbc.M113.474304
609 Laustsen, A. H., Gutierrez, J. M., Lohse, B., Rasmussen, A. R., Fernandez, J., Milbo, C., & Lomonte,
610 B. (2015). Snake venomics of monocled cobra (Naja kaouthia) and investigation of human
611 IgG response against venom toxins. *Toxicon, 99*, 23-35. doi:10.1016/j.toxicon.2015.03.001
612 Li, P., Liu, S., Lu, M., Bandyopadhyay, G., Oh, D., Imamura, T., . . . Olefsky, J. M. (2016).
613 Hematopoietic-Derived Galectin-3 Causes Cellular and Systemic Insulin Resistance. *Cell,*
614 *167*(4), 973-984 e912. doi:10.1016/j.cell.2016.10.025
615 Li, P. P., Liu, S. N., Lu, M., Bandyopadhyay, G., Oh, D., Imamura, T., . . . Olefsky, J. M. (2016).
616 Hematopoietic-Derived Galectin-3 Causes Cellular and Systemic Insulin Resistance. *Cell,*
617 *167*(4), 973-+. doi:10.1016/j.cell.2016.10.025
618 Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., . . . Adams, P. D.
619 (2019). Macromolecular structure determination using X-rays, neutrons and electrons:
620 recent developments in Phenix. *Acta Crystallogr D Struct Biol, 75*(Pt 10), 861-877.
621 doi:10.1107/S2059798319011471
622 Long, M. (2001). Evolution of novel genes. *Curr Opin Genet Dev, 11*(6), 673-680.
623 doi:10.1016/s0959-437x(00)00252-5
624 Maddux, B. A., & Goldfine, I. D. (2000). Membrane glycoprotein PC-1 inhibition of insulin receptor
625 function occurs via direct interaction with the receptor alpha-subunit. *Diabetes, 49*(1), 13-
626 19. doi:DOI 10.2337/diabetes.49.1.13
627 Margres, M. J., Rautsaw, R. M., Strickland, J. L., Mason, A. J., Schramer, T. D., Hofmann, E. P., . . .
628 Parkinson, C. L. (2021). The Tiger Rattlesnake genome reveals a complex genotype
629 underlying a simple venom phenotype. *Proc Natl Acad Sci U S A, 118*(4).
630 doi:10.1073/pnas.2014634118
631 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007).
632 Phaser crystallographic software. *J Appl Crystallogr, 40*(Pt 4), 658-674.
633 doi:10.1107/S0021889807021206
634 Mitra, J., & Bhattacharyya, D. (2014). Phosphodiesterase from Daboia russelli russelli venom:
635 purification, partial characterization and inhibition of platelet aggregation. *Toxicon, 88*, 1-
636 10. doi:10.1016/j.toxicon.2014.06.004
637 Modahl, C. M., Roointan, A., Rogers, J., Currier, K., & Mackessy, S. P. (2020). Interspecific and
638 intraspecific venom enzymatic variation among cobras (Naja sp. and Ophiophagus hannah).
639 *Comp Biochem Physiol C Toxicol Pharmacol, 232*, 108743. doi:10.1016/j.cbpc.2020.108743
640 Mori, Y., Shiratsuchi, N., Sato, N., Chaya, A., Tanimura, N., Ishikawa, S., . . . Fujita, Y. (2022).
641 Extracellular ATP facilitates cell extrusion from epithelial layers mediated by cell
642 competition or apoptosis. *Curr Biol, 32*(10), 2144-2159 e2145.
643 doi:10.1016/j.cub.2022.03.057
644 Ogawa, T., Oda-Ueda, N., Hisata, K., Nakamura, H., Chijiwa, T., Hattori, S., . . . Shibata, H. (2019).
645 Alternative mRNA Splicing in Three Venom Families Underlying a Possible Production of

Divergent Venom Proteins of the Habu Snake, Protobothrops flavoviridis. *Toxins (Basel), 11*(10). doi:10.3390/toxins11100581

Oliveira, I. S., Pucca, M. B., Wiezel, G. A., Cardoso, I. A., Bordon, K. C. F., Sartim, M. A., . . . Arantes, E. C. (2021). Unraveling the structure and function of CdcPDE: A novel phosphodiesterase from Crotalus durissus collilineatus snake venom. *Int J Biol Macromol, 178*, 180-192. doi:10.1016/j.ijbiomac.2021.02.120

Pan, C. T., & Lin, Y. S. (2020). MicroRNA retrocopies generated via L1-mediated retrotransposition in placental mammals help to reveal how their parental genes were transcribed. *Sci Rep, 10*(1), 20612. doi:10.1038/s41598-020-77381-8

Peng, C., Ren, J. L., Deng, C., Jiang, D., Wang, J., Qu, J., . . . Li, J. T. (2020). The Genome of Shaw's Sea Snake (Hydrophis curtus) Reveals Secondary Adaptation to Its Marine Environment. *Mol Biol Evol, 37*(6), 1744-1760. doi:10.1093/molbev/msaa043

Peng, L., Xu, X., Shen, D., Zhang, Y., Song, J., Yan, X., & Guo, M. (2011). Purification and partial characterization of a novel phosphodiesterase from the venom of Trimeresurus stejnegeri: inhibition of platelet aggregation. *Biochimie, 93*(9), 1601-1609. doi:10.1016/j.biochi.2011.05.027

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem, 25*(13), 1605-1612. doi:10.1002/jcc.20084

Portillo, F., Stanley, E. L., Branch, W. R., Conradie, W., Rodel, M. O., Penner, J., . . . Greenbaum, E. (2019). Evolutionary history of burrowing asps (Lamprophiidae: Atractaspidinae) with emphasis on fang evolution and prey selection. *PLoS One, 14*(4), e0214889. doi:10.1371/journal.pone.0214889

Rao, W. Q., Kalogeropoulos, K., Allentoft, M. E., Gopalakrishnan, S., Zhao, W. N., Workman, C. T., . . . Laustsen, A. H. (2022). The rise of genomics in snake venom research: recent advances and future perspectives. *Gigascience, 11*. doi:10.1093/gigascience/giac024

Reyes-Velasco, J., Card, D. C., Andrew, A. L., Shaney, K. J., Adams, R. H., Schield, D. R., . . . Castoe, T. A. (2015). Expression of venom gene homologs in diverse python tissues suggests a new model for the evolution of snake venom. *Mol Biol Evol, 32*(1), 173-183. doi:10.1093/molbev/msu294

Santoro, M. L., Vaquero, T. S., Paes Leme, A. F., & Serrano, S. M. (2009). NPP-BJ, a nucleotide pyrophosphatase/phosphodiesterase from Bothrops jararaca snake venom, inhibits platelet aggregation. *Toxicon, 54*(4), 499-512. doi:10.1016/j.toxicon.2009.05.016

Sloutskin, A., Danino, Y. M., Orenstein, Y., Zehavi, Y., Doniger, T., Shamir, R., & Juven-Gershon, T. (2015). ElemeNT: a computational tool for detecting core promoter elements. *Transcription, 6*(3), 41-50. doi:10.1080/21541264.2015.1067286

Sorek, R. (2007). The birth of new exons: mechanisms and evolutionary consequences. *RNA, 13*(10), 1603-1608. doi:10.1261/rna.682507

Stock, R. P., Massougbodji, A., Alagon, A., & Chippaux, J. P. (2007). Bringing antivenoms to Sub-Saharan Africa. *Nat Biotechnol, 25*(2), 173-177. doi:10.1038/nbt0207-173

686 Storer, J., Hubley, R., Rosen, J., Wheeler, T. J., & Smit, A. F. (2021). The Dfam community resource of
687 transposable element families, sequence models, and genome annotations. *Mob DNA,*
688 *12*(1), 2. doi:10.1186/s13100-020-00230-y
689 Suryamohan, K., Krishnankutty, S. P., Guillory, J., Jevit, M., Schroder, M. S., Wu, M., . . . Seshagiri, S.
690 (2020). The Indian cobra reference genome and transcriptome enables comprehensive
691 identification of venom toxins. *Nat Genet, 52*(1), 106-117. doi:10.1038/s41588-019-0559-8
692 Tan, C. H., Tan, K. Y., Fung, S. Y., & Tan, N. H. (2015). Venom-gland transcriptome and venom
693 proteome of the Malaysian king cobra (Ophiophagus hannah). *BMC Genomics, 16*, 687.
694 doi:10.1186/s12864-015-1828-2
695 Tan, N. H., & Tan, C. S. (1988). A comparative study of cobra (Naja) venom enzymes. *Comp Biochem*
696 *Physiol B, 90*(4), 745-750. doi:10.1016/0305-0491(88)90329-x
697 Trummal, K., Aaspollu, A., Tonismagi, K., Samel, M., Subbi, J., Siigur, J., & Siigur, E. (2014).
698 Phosphodiesterase from Vipera lebetina venom - structure and characterization. *Biochimie,*
699 *106*, 48-55. doi:10.1016/j.biochi.2014.07.020
700 Tsai, S. H., Kinoshita, M., Kusu, T., Kayama, H., Okumura, R., Ikeda, K., . . . Takeda, K. (2015). The
701 Ectoenzyme E-NPP3 Negatively Regulates ATP-Dependent Chronic Allergic Responses by
702 Basophils and Mast Cells. *Immunity, 42*(2), 279-293. doi:10.1016/j.immuni.2015.01.015
703 Tsetlin, V. I. (2015). Three-finger snake neurotoxins and Ly6 proteins targeting nicotinic
704 acetylcholine receptors: pharmacological tools and endogenous modulators. *Trends*
705 *Pharmacol Sci, 36*(2), 109-123. doi:10.1016/j.tips.2014.11.003
706 Valerio, A. A., Corradini, A. C., Panunto, P. C., Mello, S. M., & Hyslop, S. (2002). Purification and
707 characterization of a phosphodiesterase from Bothrops alternatus snake venom. *J Protein*
708 *Chem, 21*(8), 495-503. doi:10.1023/a:1022414503995
709 Verta, J. P., & Jacobs, A. (2022). The role of alternative splicing in adaptation and evolution. *Trends*
710 *Ecol Evol, 37*(4), 299-308. doi:10.1016/j.tree.2021.11.010
711 Westeen, E. P., Durso, A. M., Grundler, M. C., Rabosky, D. L., & Davis Rabosky, A. R. (2020). What
712 makes a fang? Phylogenetic and ecological controls on tooth evolution in rear-fanged
713 snakes. *BMC Evol Biol, 20*(1), 80. doi:10.1186/s12862-020-01645-0
714 Wheeler, T. J., & Eddy, S. R. (2013). nhmmer: DNA homology search with profile HMMs.
715 *Bioinformatics, 29*(19), 2487-2489. doi:10.1093/bioinformatics/btt403
716

717

Figures

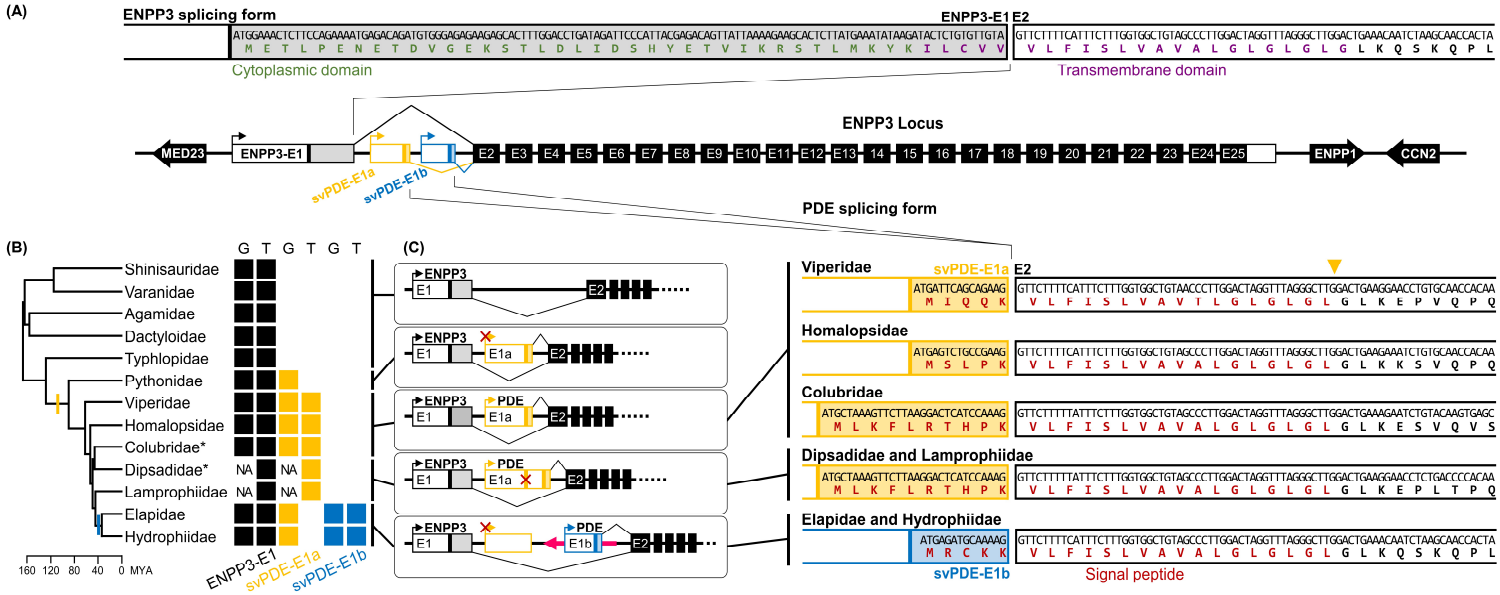


Fig. 1 (A) N-terminal of peptide sequences of ENPP3 and PDE encoded by alternative splicing. The scheme of exons and genomic synteny are presented with flanking coding genes. The 5' untranslated regions (5' UTR) and coding regions of first exons are separated by colored vertical lines, which are the translation start codons. Only the lengths of E1, E1a and E1b are drawn in scale. Putative promoters identified in the proximal upstream are indicated with arrows attached to the first exons. The same elements are expressed by the same colors throughout the figure. The sequences around splicing junctions of E1 and E2 are zoomed in with translated peptides of functional domains highlighted in corresponding colors. The yellow triangle indicates the putative cleavage site of the signal peptidase. (B) Presences and absences of ENPP3-specific and svPDE-specific E1s in different clades. Color-filled squares denote the exons identified in the genomes "G" and transcriptomes "T", where "NA" indicates the clades without available genome assemblies. Clades with asterisks use Duvernoy's glands as their venom delivery system. The colored vertical lines on the phylogenetic tree indicate the inferred branches that svPDE-E1a (yellow) and svPDE-E1b (blue) emerged. (C) The scheme of alternative splicing of ENPP3 and svPDE transcripts in different clades. The possible malfunctioned elements, including svPDE-E1a promoters in Pythonidae, Elapidae and Hydrophiidae, and ancestral translation start sites in Dipsadidae and Lamprophiidae, are crossed out in red. The long pink arrow indicates the mobile element.

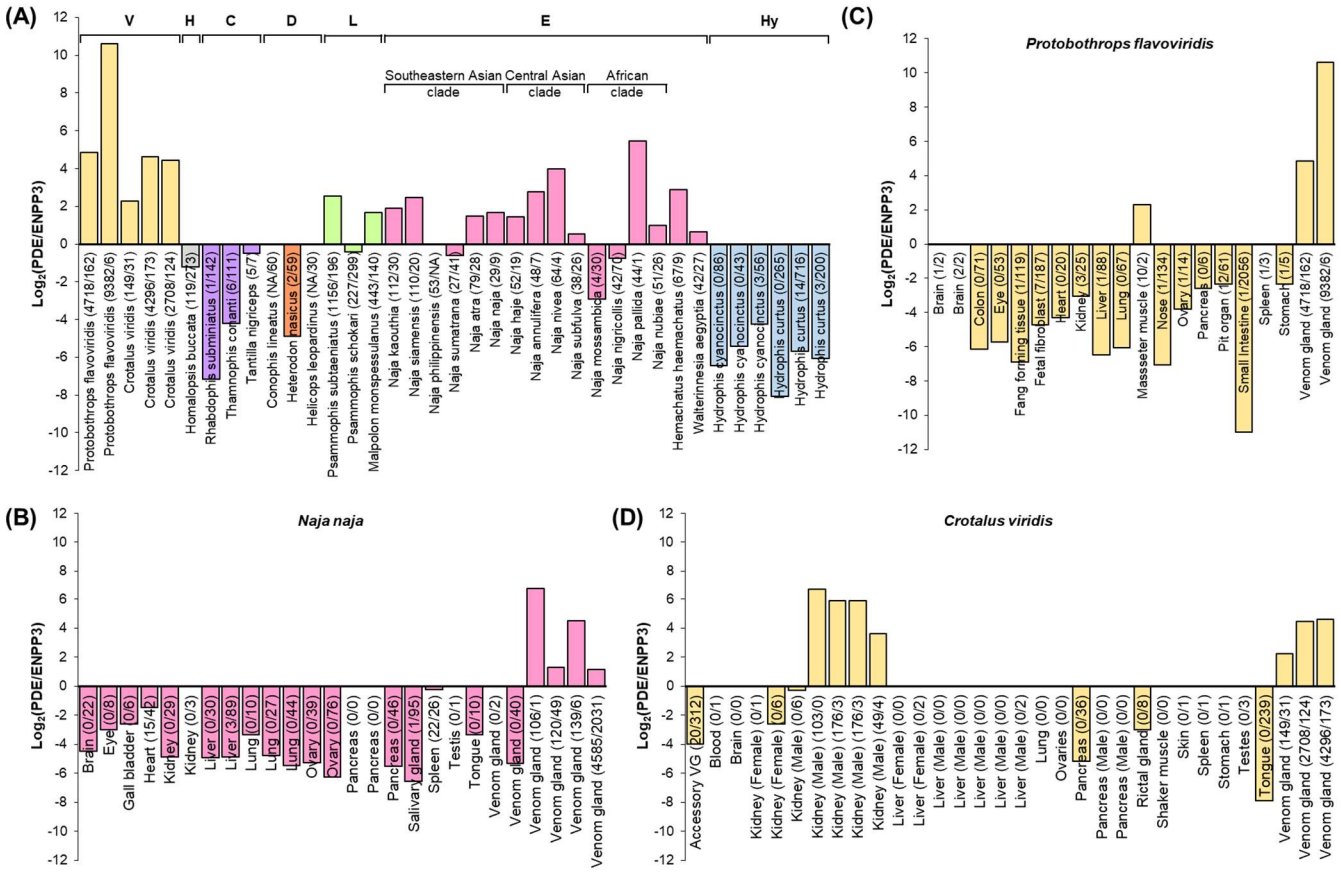


Fig. 2 Species-specific and tissue-specific expression of svPDE over ENPP3 splicing forms. The expression of each splicing form was calculated by the number of reads spanning the specific E1s and the conserved E2. The log2 transformed ratios are shown in the figure and a count of zero was adjusted to a pseudo count of one for valid log transformation. The unadjusted counts are presented in the axis labels as the format of svPDE/ENPP3. The alternative splicing forms not identified in the transcriptomes and no available genomes to infer their presence are denoted as NA. (A) Species-specific expressions in venom glands and, for Colubridae and Dipsadidae, Duvernoy's gland, an anatomic gland structure similar to venom glands, were compared. Samples with less than five svPDE transcripts and less than five ENPP3 transcripts are regarded as having no expression for both transcripts and not shown. Clades are colored in yellow, grey, purple, orange, green, pink and blue with the clade prefixes of Viperidae, Homalopsidae, Colubridae, Dipsadidae, Lamprophiidae, Elapidae and Hydrophiidae, respectively. (B-D) Tissue-specific expression of svPDE over ENPP3 in selected species: (B) *Naja naja* (Elapidae) and (C) *Protobothrops flavoviridis* (Viperidae) and (D) *Crotalus viridis* (Viperidae).

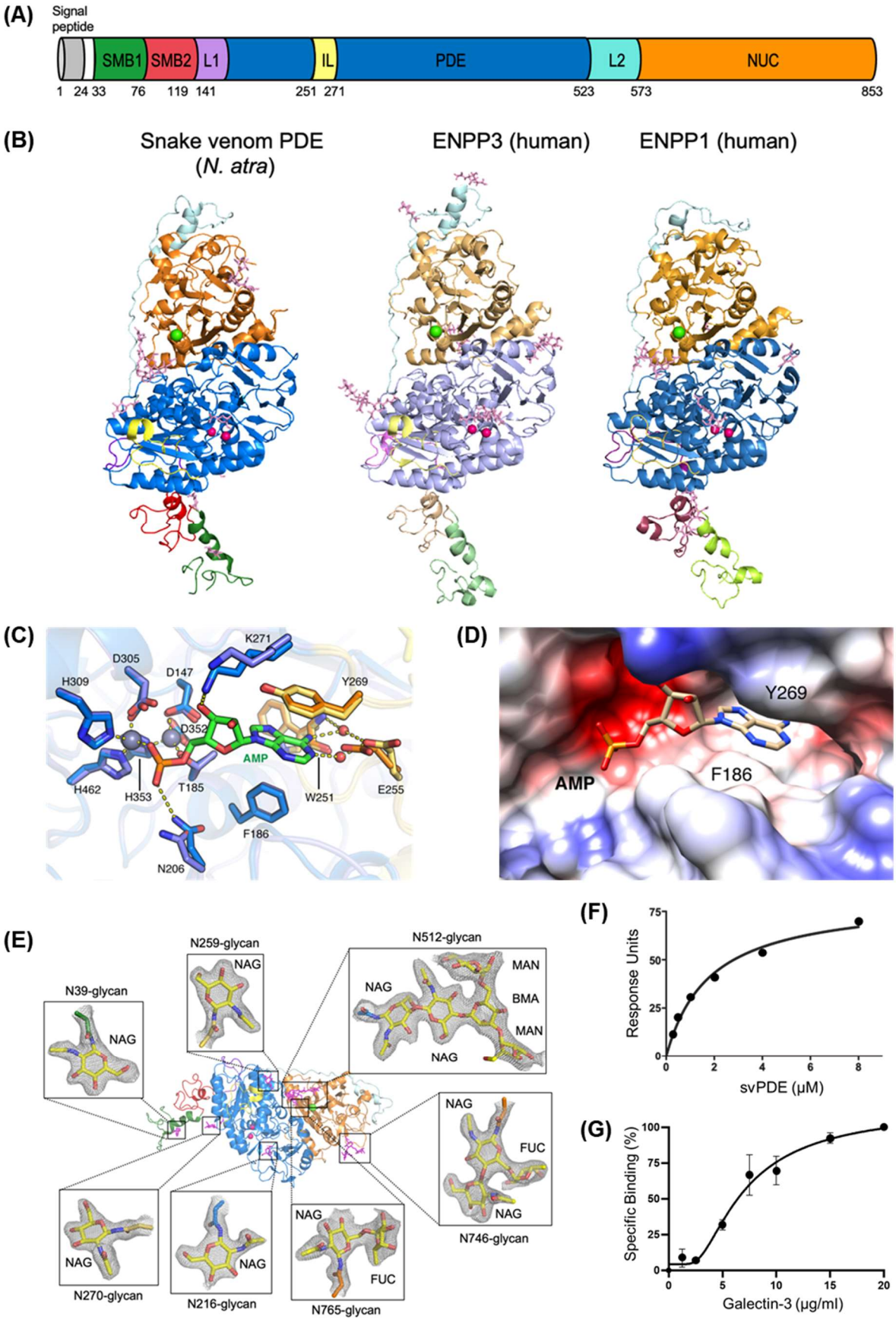


Fig. 3 (A) Domain organizations of *Naja atra* svPDE. SMB1, somatomedin-B-like 1 domain; SMB2, somatomedin-B-like 2 domain; L1, loop 1; PDE, catalytic phosphodiesterase domain; IL, insertion loop; L2, loop 2; NUC, nuclease-like domain. (B) Crystal structures of svPDE from *Naja atra* (PDB code: 5GZ4), human ENPP3 (PDB code: 6C01) and human ENPP1 (PDB code: 6WET) in cartoon representation. Zinc and calcium ions are shown as hot pink and green spheres, respectively. N-glycans are shown as light pink sticks. (C) Superposition of active sites of svPDE from *Naja atra* in the apo and AMP-complexed forms. Zinc atoms are shown as grey spheres and water atoms are shown as red ones. Residues involving AMP binding in the apo and AMP-complexed forms are shown in slate/orange and marine/yellow, respectively. Hydrogen bonds and coordinate bonds are shown as dashed yellow lines. (D) Electrostatic potential surface of the nucleotide-binding pocket of svPDE from *Naja atra* in the AMP-complexed form. (E) N-glycans at Asn39, Asn216, Asn259, Asn270, Asn512 and Asn746 of svPDE from *Naja atra* are shown as sticks with electron densities. 2Fo-Fc electron density maps contoured at 1.0σ. (F) Surface plasmon resonance (SPR) investigation of the binding between svPDE (0.25, 0.5, 1.0, 2.0, 4.0, and 8.0 μM) and the immobilized insulin receptor. K_D of svPDE binding to the insulin receptor were obtained by steady-state affinity model. (G) Binding affinity measurement of Gal-3 with svPDE from *Naja atra*. Standard deviations of three replicates were indicated.

Supporting Information for

The evolution and structure of snake venom phosphodiesterase (svPDE) highlight its importance in venom actions

Cheng-Tsung Pan[1,2], Chien-Chu Lin[1], I-Jin Lin[1], Kun-Yi Chien[3], Yeong-Shin Lin[4]*, Hsiao-Han Chang[1,5,6]*, Wen-Guey Wu[1,5]*

[1] Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, Taiwan

[2] Department of Biostatistics, University of Oslo, Oslo, Norway

[3] Graduate Institute of Biomedical Sciences, Department of Biochemistry and Molecular Biology, School of Medicine, Chang Gung University, Taiwan

[4] Institute of Bioinformatics and Systems Biology, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

[5] Department of Life Science, National Tsing Hua University, Hsinchu, Taiwan

[6] Institute of Molecular and Cellular Biology, National Tsing Hua University, Hsinchu, Taiwan

*Co-corresponding authors: Yeong-Shin Lin, Hsiao-Han Chang, Wen-Guey Wu

Email: YSL yslin@nycu.edu.tw

HHC hhchang@life.nthu.edu.tw

WGW wgwu@life.nthu.edu.tw

This PDF file includes:

Figures S1 to S9

Tables S1 to S6

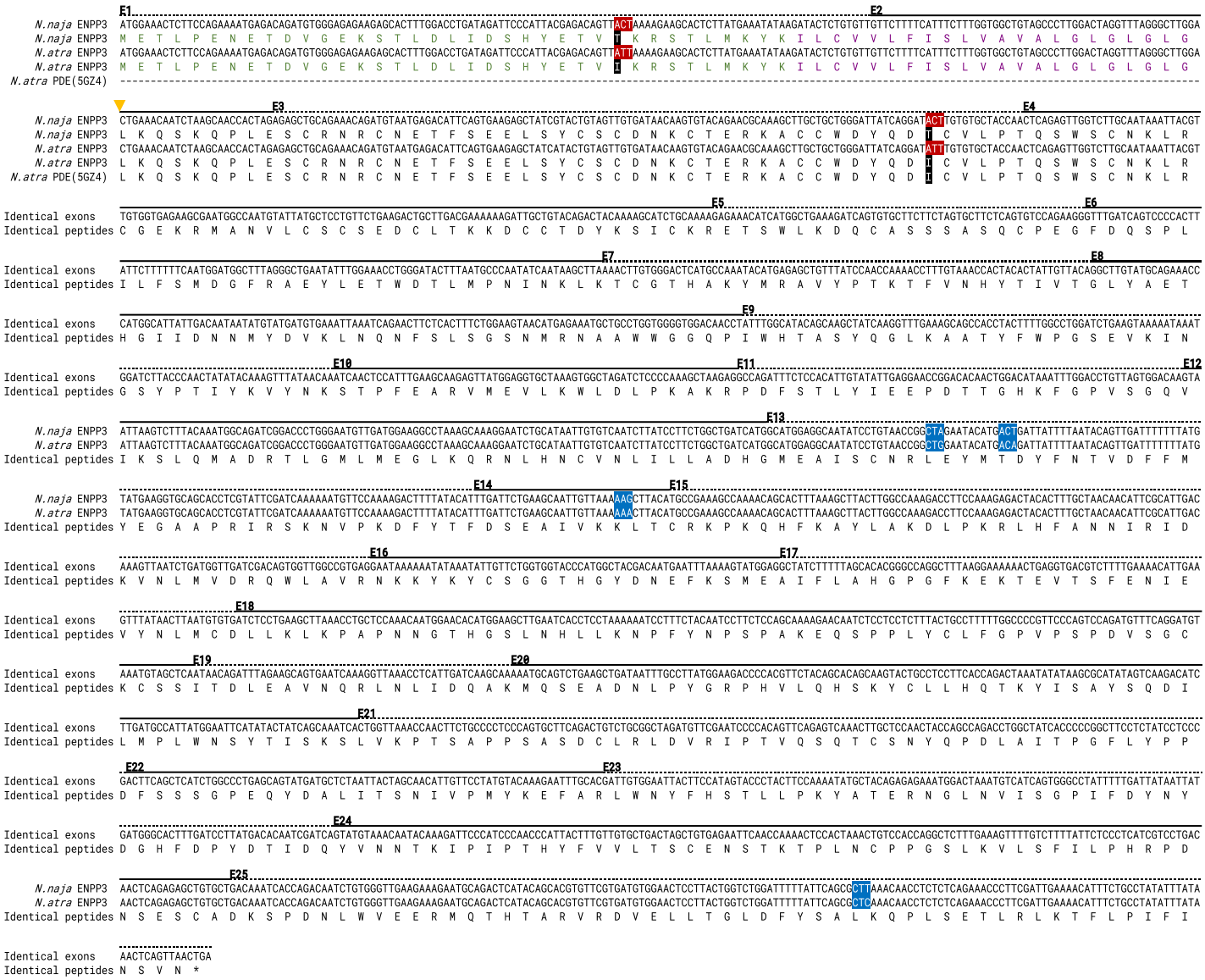


Fig. S1 The sequence alignment of the translated peptides of ENPP3 inferred from the genomic exons (*N. naja* and *N. atra*) and the svPDE peptides revealed from proteomic approaches. Regions of each exon are underlined with solid and dotted lines alternately. The cytoplasmic and transmembrane domains are colored in green and plum, respectively. The codons with non-synonymous substitutions and the corresponding amino acids are highlighted in red and black, respectively. The codons with synonymous substitutions are highlighted in blue. Identical sequences are collapsed into one sequence for representation. The yellow arrow indicates the putative cleavage site of the signal peptide, which is in line with the N-terminal end of the purified svPDE protein.

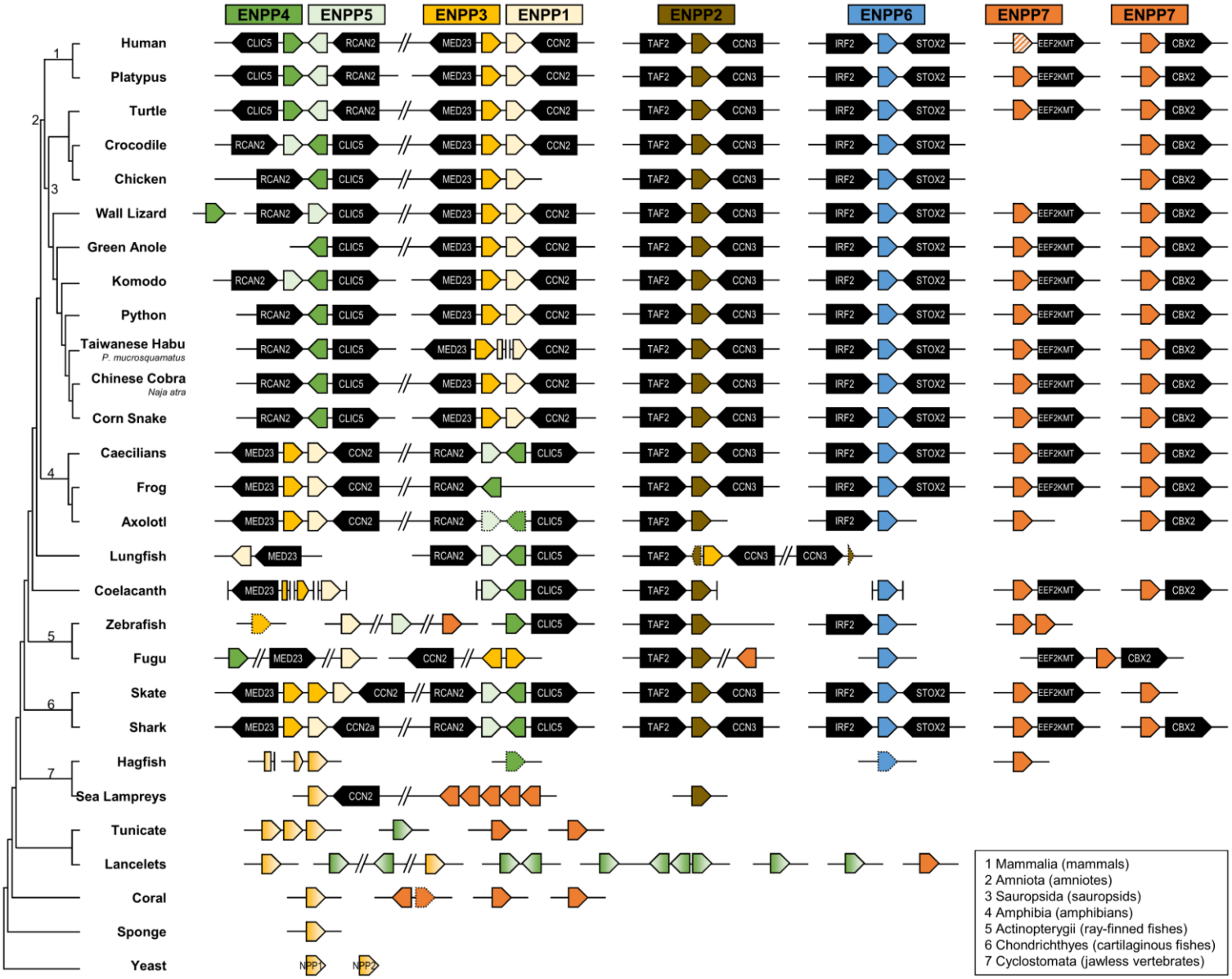
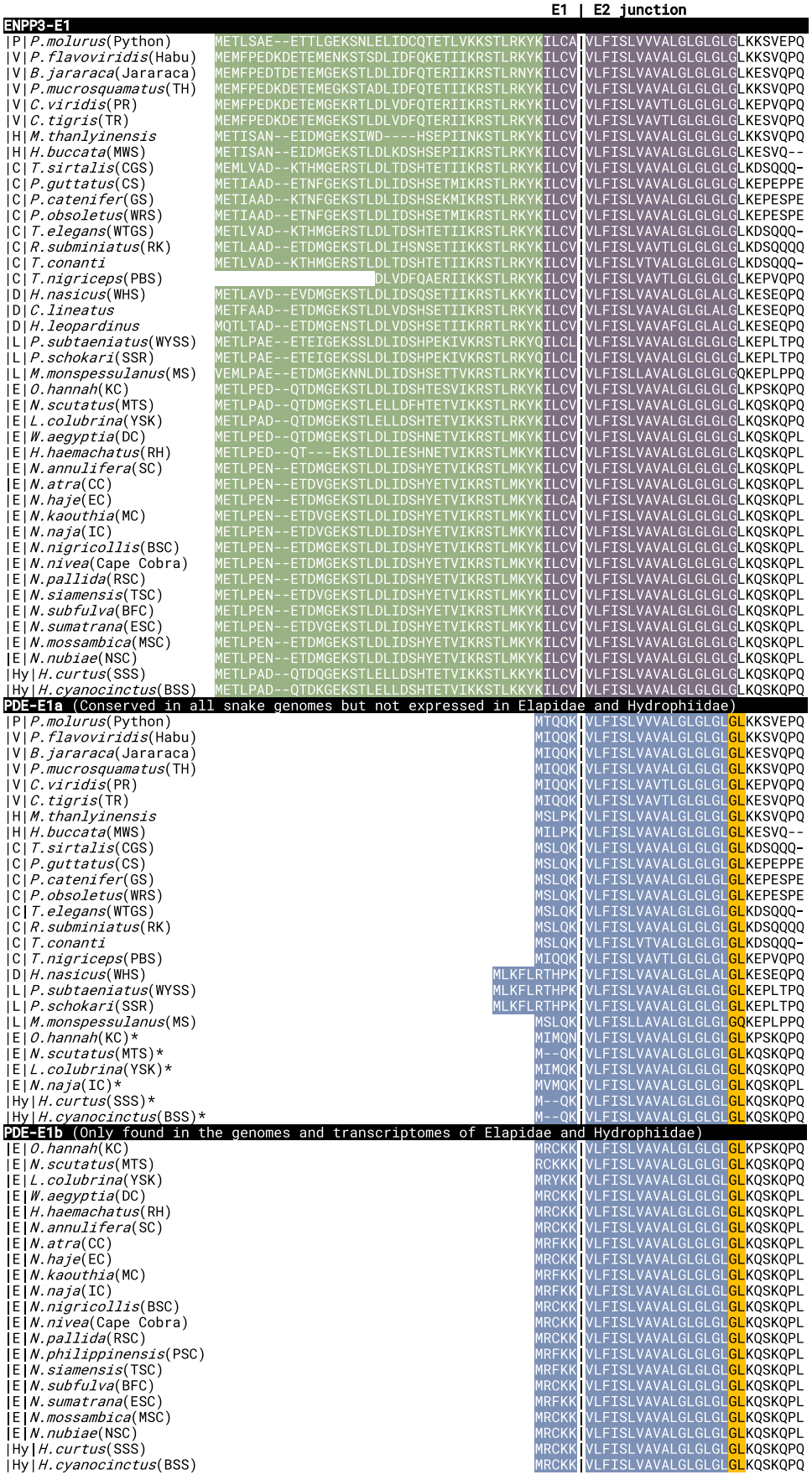


Fig. S2 Presences and absences of ENPP family members across eukaryotes. Seven members of ENPP families are represented in pentagons with corresponding colors. Their genomic synteny are shown with the conserved flanking genes (blacked pentagons). ENPP genes with gradient colors are genes diverged before the divergence of two named ENPP family members (e.g., a gene phylogenetically clustered with the common ancestor of ENPP3 and ENPP1 is drawn in a gradient yellow). Genes with only a few conserved exons regarded as evolutionary relics are indicated with dotted borders. One human ENPP7 denoted by slashes is a confirmed pseudogene in the latest genome annotation release. Genes on a single continuous horizontal line are located on the same contig or chromosome. Vertical lines indicate the assembled ends of contigs or chromosomes. Genes fragmentedly located on different contigs that could infer the coordinates by genomic synteny are drawn by order. The branch lengths of the cladogram are not drawn proportional to the evolutionary time.

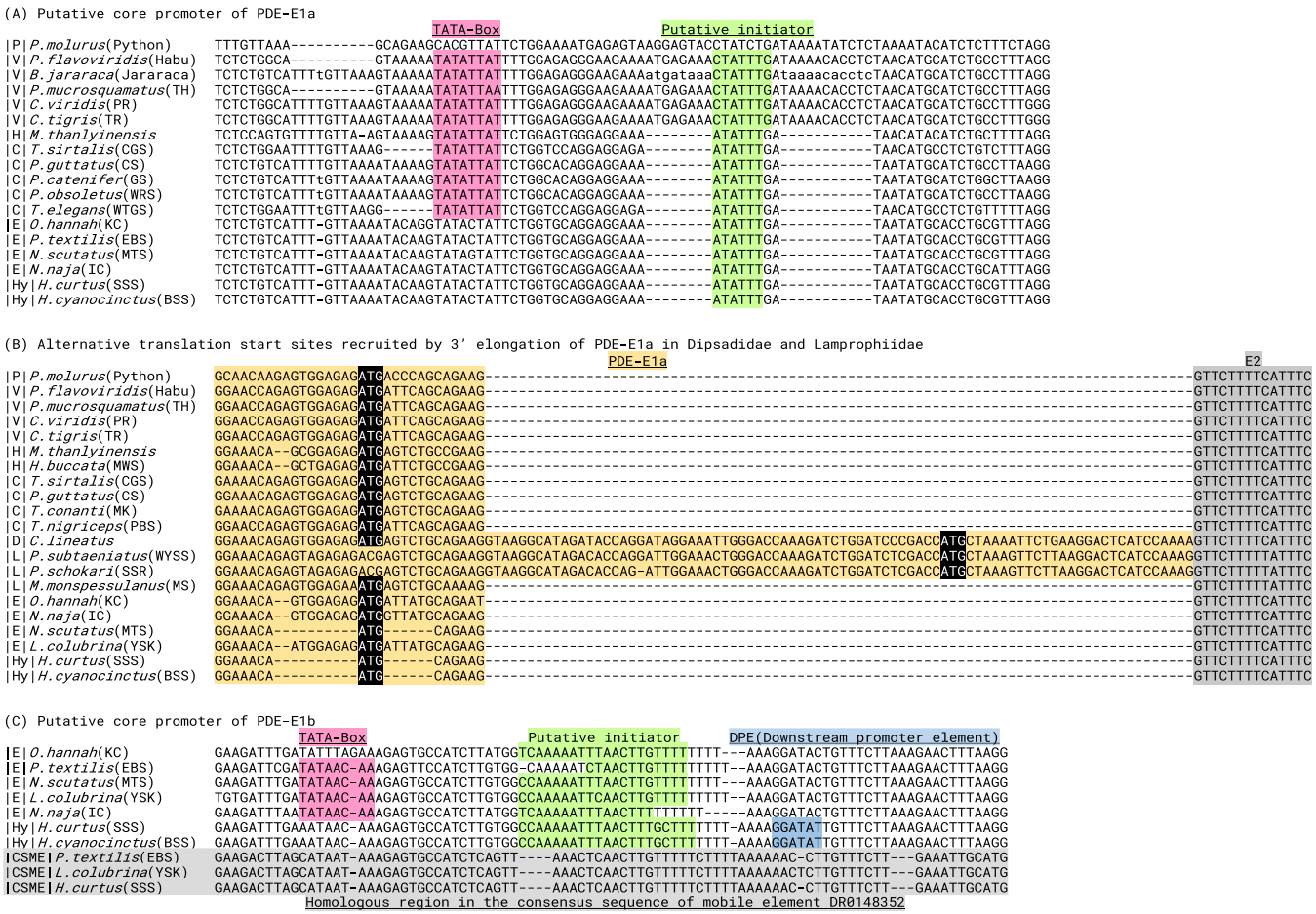


* Conserved in the genomes but not expressed in the transcriptomes (only selected sequences were shown).

Abbreviations for the common names:
Taiwanese habu (TH), Prairie rattlesnake (PR), Terrestrial rattlesnake (TR), Masked water snake (MWS), Common garter snake (CGS), Corn snake (CS), Gopher snake (GS), Western rat snake (WRS), Western terrestrial garter snake (WTGS), Red-necked keelback (RK), Plain black-headed snake (PBS), Western hognose snake (WHS), Western yellow-bellied sand snake (WYSS), Schokari sand racer (SSR), Montpellier snake (MS), King cobra (KC), Mainland tiger snake (MTS), Yellow-lipped sea krait (YSK), India cobra (IC), Desert cobra (DC), Rinkhals (RH), Snouted cobra (SC), Chinese cobra (CC), Egyptian cobra (EC), Monocled cobra (MC), Black-necked spitting cobra (BSC), Red spitting cobra (RSC), Philippine spitting cobra (PSC), Thai spitting cobra (TSC), Brown forest cobra (BFC), Equatorial spitting cobra (ESC), Mozambique spitting cobra (MSC), Nubian spitting cobra (NSC), Shaw's sea snake (SSS), Blue-banded sea snake (BSS).

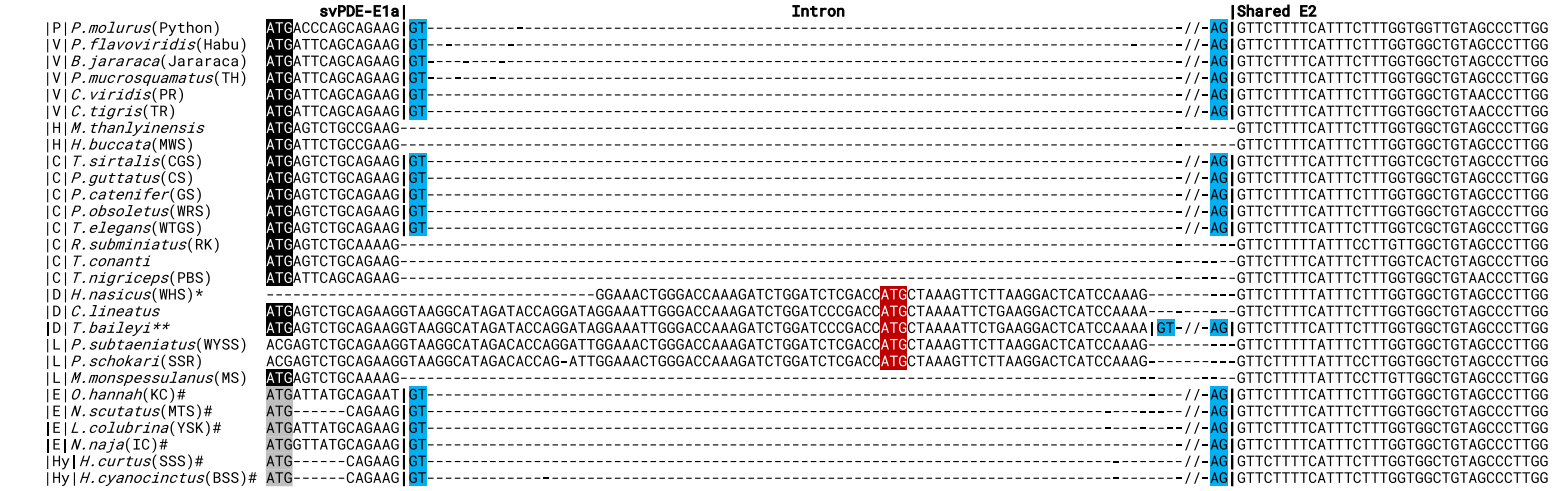
Abbreviations for the clades:
Pythonidae (P), Viperidae (V), Homalopsidae (H), Colubridae (C), Dipsadidae (D), Lamprophiidae (L), Elapidae (E) and Hydrophiidae (Hy).

Fig. S3 The multiple sequence alignment of the translated peptides around the E1-E2 junction region. Peptides started from the start codon (ATG, Methionine) of exon 1 to the 5' partial region of exon 2 are shown. The cytoplasmic domains, the transmembrane domains, signal peptides and cleavage sites are colored in green, plum, blue and yellow. Abbreviations are listed in the figure. Only a few species are selected as representatives of Elapidae and Hydrophiidae, whose svPDE-E1a is conserved in genomes but not expressed in the transcriptomes. The ENPP3 transcript of *T. nigriceps* was partially assembled and does not cover the 5' end of the coding sequence.



Abbreviations
TH: Taiwanese habu, PR: Prairie rattlesnake, TR: Terrestrial rattlesnake, CGS: Common garter snake, CS: Corn snake, GS: Gopher snake, WRS: Western rat snake, WTGS: Western terrestrial garter snake, MK: Milk snake, PBS: Plain black-headed snake, WYSS: Western yellow-bellied sand snake, SSR: Schokari sand racer, MS: Montpellier snake, KC: King cobra, EBS: Eastern brown snake, MTS: Mainland tiger snake, YSK: Yellow-lipped sea krait, IC: India cobra, SSS: Shaw's sea snake and BSS: Blue-banded sea snake. P: Pythonidae, V: Viperidae, H: Homalopsidae, C: Colubridae, E: Elapidae, Hy: Hydrophiidae and CSME: Consensus sequence of mobile element DR0148352.

Fig. S4 The multiple sequence alignment of the putative core promoters located on the upstream of (A) svPDE-E1a and (B) svPDE-E1b. Putative functional motifs are highlighted with corresponding colors. Abbreviations of clades and species common names are the same as Supplementary fig. S3 with an additional one, CSME, as the consensus sequence of mobile element (DR0148352).



* Missing of 5'end during sequencing was probably due to the nature of poly-A library preparation.
** The most closely related species to *C.lineatus* with available genome.
\# These sequences are conserved in the genomes but not expressed in the transcriptomes.

Abbreviations for the common names:
Taiwanese habu (TH), Prairie rattlesnake (PR), Terrestrial rattlesnake (TR), Masked water snake (MWS), Common garter snake (CGS), Corn snake (CS), Gopher snake (GS), Western rat snake (WRS), Western terrestrial garter snake (WTGS), Red-necked keelback (RK), Plain black-headed snake (PBS), Western hognose snake (WHS), Western yellow-bellied sand snake (WYSS), Schokari sand racer (SSR), Montpellier snake (MS), King cobra (KC), Mainland tiger snake (MTS), Yellow-lipped sea krait (YSK), India cobra (IC), Shaw's sea snake (SSS), Blue-banded sea snake (BSS), Philippine spitting cobra (PSC).

Abbreviations for the clades:
Pythonidae (P), Viperidae (V), Homalopsidae (H), Colubridae (C), Dipsadidae (D), Lamprophiidae (L), Elapidae (E) and Hydrophiidae (Hy).

Fig. S5 The multiple sequence alignment of the coding regions of svPDE-E1a and 5' partial E2. Conserved start codons of svPDE-E1a are highlighted in black and gray, corresponding to the expressed and non-expressed transcripts, respectively. For all species with available genomes, the interval sequences between svPDE-E1a and conserved E2 are canonical GT-AG introns, which are shown in the alignments with splicing sites highlighted in blue. Other regions of intron sequences are omitted for a clear view. For species that only have transcriptomic data, the sequences shown in the alignment were obtained from the assembled transcripts. In Dipsadidae and Lamprophiidae, the alternative start codons located on the elongated svPDE-1a due to using alternative 3' splicing sites are highlighted in red.

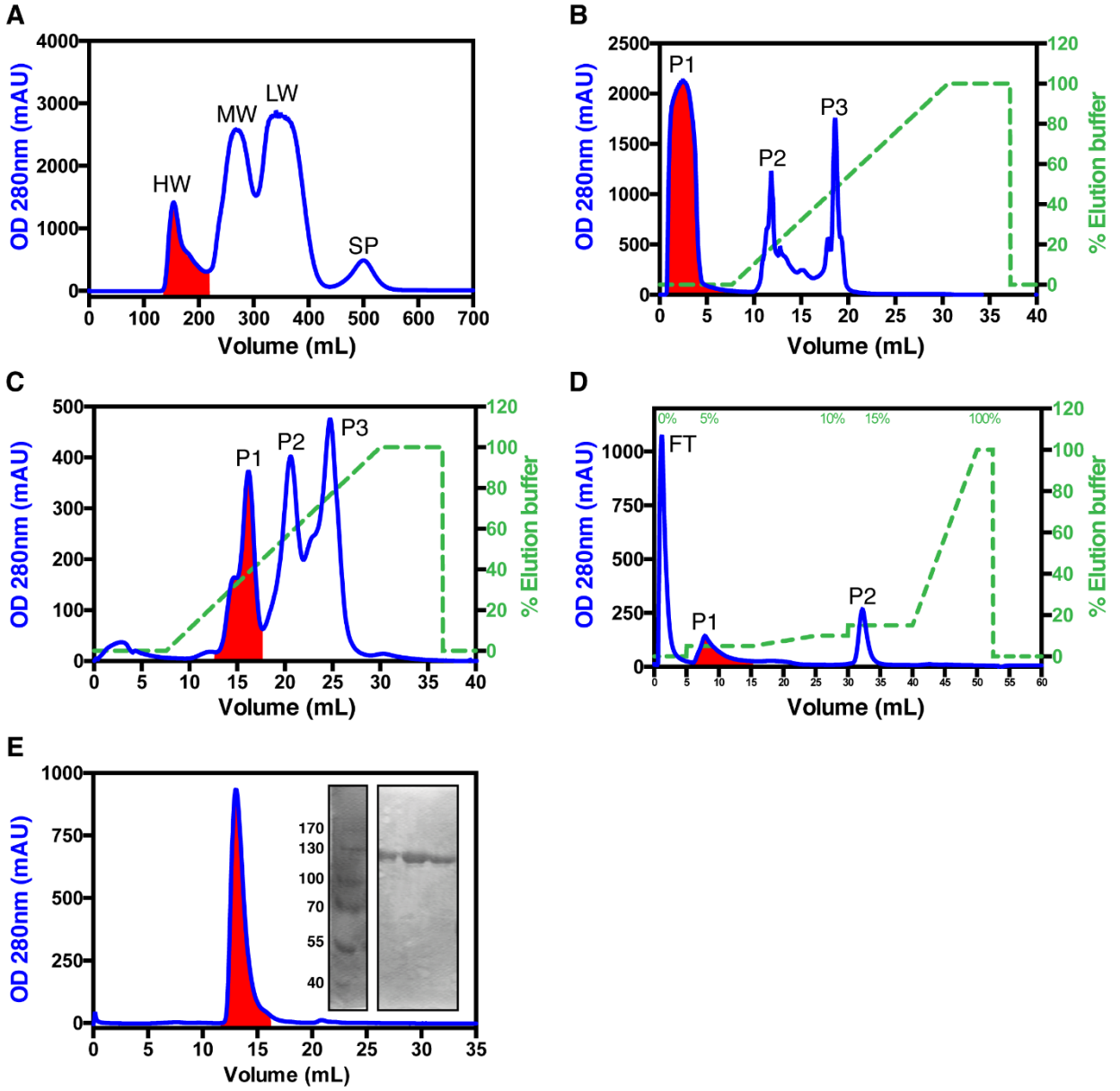


Fig. S6 Purification of svPDE from the crude venom of *Naja atra* habitated in Taiwan. SvPDE purification was performed using (A) Sephadex G-75 followed by (B) MonoQ, (C) Hitrap Heparin, (D) MonoS, and (E) Superdex 200 columns. Red-highlighted fractions contained svPDE. (A) At first, size exclusion chromatography separated the crude venom into four fractions: high (HW), medium (MW), and low molecular weight proteins (LW), and small peptides (SP), in which svPDE with a molecular weight of ~100k Da appeared in the HW, the first fraction of the elution. (B) Secondly, anion-exchange chromatography with a MonoQ column captured negatively charged proteins, leaving svPDE in P1, the flow-through fraction. (C) The next affinity chromatography (HiTrap Heparin) and (D) cation-exchange chromatography (MonoS) further separated svPDE from other high molecular weight venom components. (E) Lastly, the ultra-pure svPDE was acquired after the Superdex 200 column. The purity of svPDE analyzed by SDS-PAGE under reducing conditions is shown in the inset of the figure (see Fig. S6 source data for original uncropped image). The svPDE purified from *Naja atra* venom was estimated to be about 0.1-0.2% of crude venom.

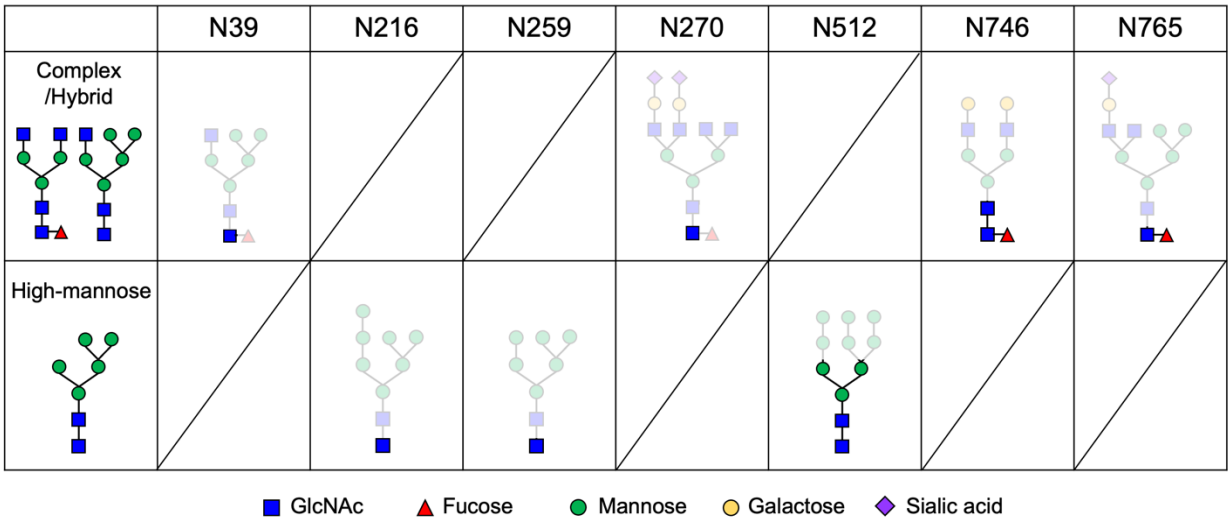


Fig. S7 Identification of N-glycosylated sites and N-glycan patterns of svPDE from *Naja atra* based on electron density distributions from X-ray diffraction data and mass spectrometric methods. The most complete composition of the major form at each site is depicted in the table. Glycan residues observed in crystal structures are highlighted in dark colors and those depicted as light colors were only observed according to the mass spectrometric data. Four glycans (N39, N270, N746 and N765) were fucosylated and three glycans (N216, N259 and N512) adopted high mannose structures.

The raw data can be retrieved from the following links:

https://docs.google.com/spreadsheets/d/1kEU9ofVK0J1IopM55m4a1Rd7PgSMlblu/edit?usp=sharing&ouid=114353426449521939124&rtpof=true&sd=true

https://docs.google.com/spreadsheets/d/145r7S9KRE62ZrqPVcfWQODdY532WbhDf/edit?usp=sharing&ouid=114353426449521939124&rtpof=true&sd=true

https://docs.google.com/spreadsheets/d/1vM9Va4xntFPiTIlz9YDFV2RHe4EQjT3-/edit?usp=sharing&ouid=114353426449521939124&rtpof=true&sd=true

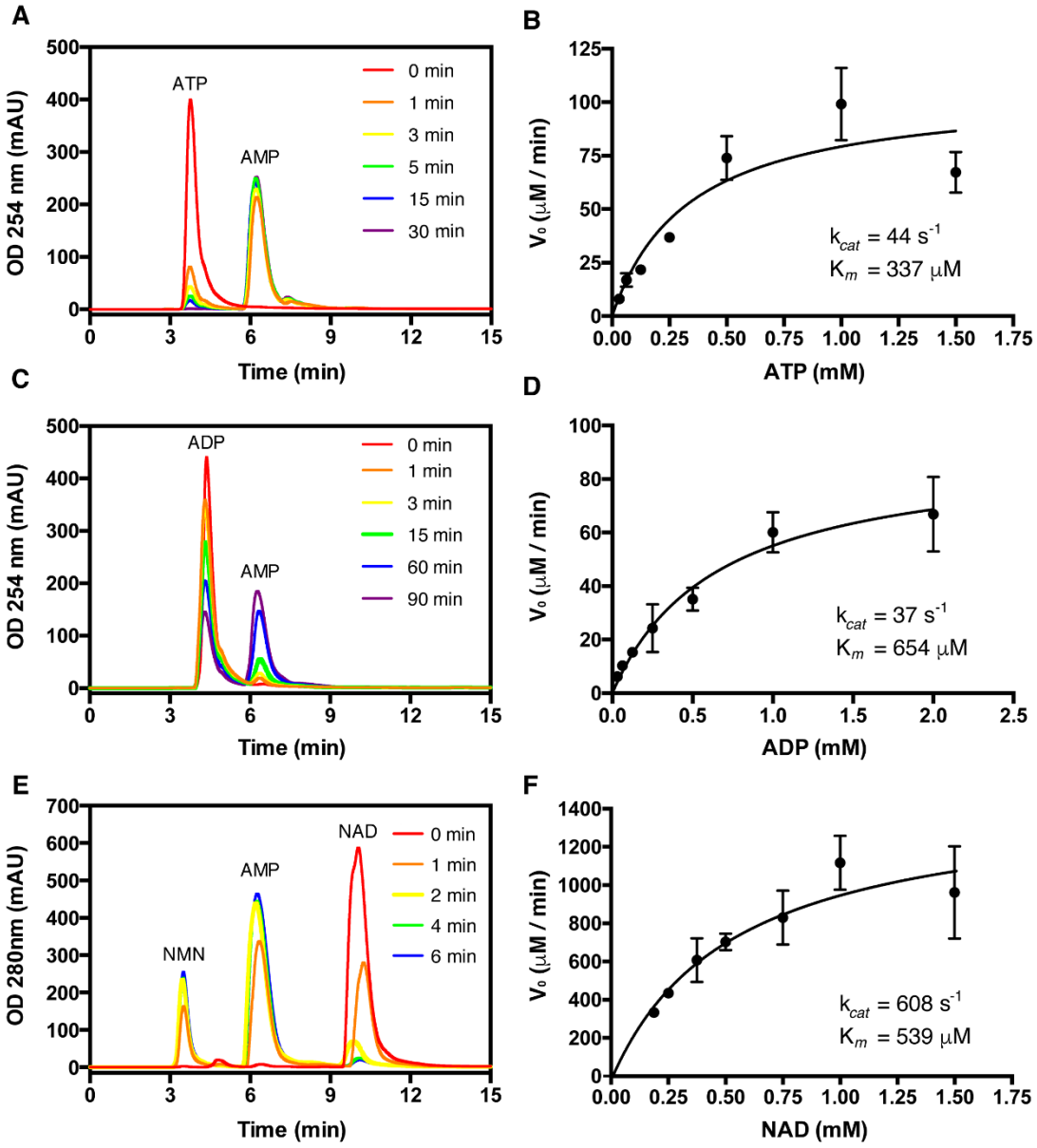


Fig. S8 Nucleotide and NAD-hydrolysis activities of svPDE from *Naja atra*. (A) HPLC analysis of 40 nM svPDE incubated with 1 mM ATP for 0, 1, 3, 5, 15 and 30 min (AU: absorbance unit). (B) Michaelis-Menten plot of the hydrolysis of ATP by 40 nM svPDE. (C) HPLC analysis of 40 nM svPDE incubated with 1 mM ADP for 0, 1, 3, 5, 15, 60 and 90 min. (D) Michaelis-Menten plot of the hydrolysis of ADP by 40 nM svPDE. (E) HPLC analysis of 80 nM svPDE incubated with 1 mM NAD for 0, 1, 2, 4 and 6 min. (F) Michaelis-Menten plot of the hydrolysis of NAD by 80 nM svPDE (NMN: nicotinamide mononucleotide).

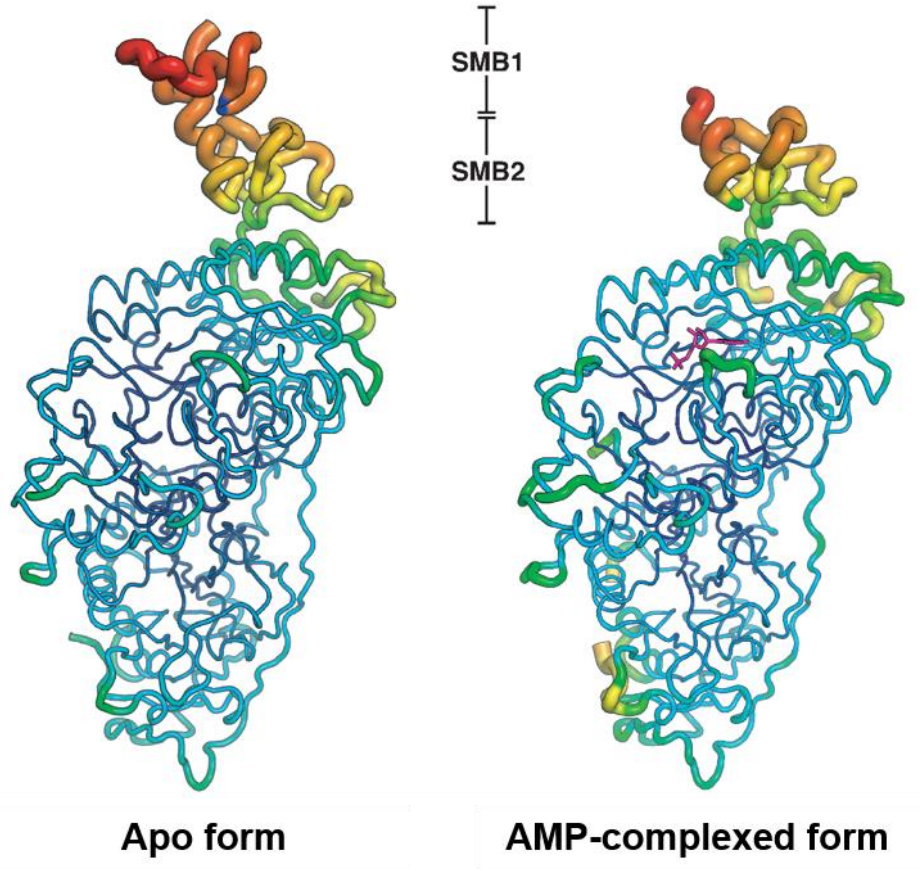


Fig. S9 Putty (sausage) representation of the crystal structures of svPDE from Naja atra. The radii of the backboned structures are proportional to the averaged B-factors of the individual residues (or the C_α atoms), which are color-ramped from blue to red for low to high B-factors.

Table S1 Summary of genomic and transcriptomic data used in this study

Family	NCBI accession	Data type	Scientific name	Common name	Source Tissue for RNA-Seq
Shinisauridae	GCA_021292165.1	Genome	*Shinisaurus crocodilurus*	Chinese crocodile lizard	
Shinisauridae	SRR14583706	RNA-Seq	*Shinisaurus crocodilurus*	Chinese crocodile lizard	Mixture
Varanidae	GCF_004798865.1	Genome	*Varanus komodoensis*	Komodo dragon	
Varanidae	SRR8466822	RNA-Seq	*Varanus komodoensis*	Komodo dragon	blood
Varanidae	SRR8466824	RNA-Seq	*Varanus komodoensis*	Komodo dragon	blood
Varanidae	SRR8466825	RNA-Seq	*Varanus komodoensis*	Komodo dragon	blood
Varanidae	SRR8466826	RNA-Seq	*Varanus komodoensis*	Komodo dragon	blood
Varanidae	SRR8466827	RNA-Seq	*Varanus komodoensis*	Komodo dragon	blood
Varanidae	SRR8735151	RNA-Seq	*Varanus komodoensis*	Komodo dragon	Heart
Varanidae	SRR8735152	RNA-Seq	*Varanus komodoensis*	Komodo dragon	Heart
Agamidae	GCF_900067755.1	Genome	*Pogona vitticeps*	Central bearded dragon	
Agamidae	SRR14455628	RNA-Seq	*Pogona vitticeps*	Central bearded dragon	Epithelium from dental tissues
Agamidae	SRR14455629	RNA-Seq	*Pogona vitticeps*	Central bearded dragon	Epithelium from dental tissues
Agamidae	SRR14455630	RNA-Seq	*Pogona vitticeps*	Central bearded dragon	Epithelium from dental tissues
Agamidae	SRR14455631	RNA-Seq	*Pogona vitticeps*	Central bearded dragon	Mesenchyme from dental tissues
Agamidae	SRR14455632	RNA-Seq	*Pogona vitticeps*	Central bearded dragon	Mesenchyme from dental tissues
Agamidae	SRR14455633	RNA-Seq	*Pogona vitticeps*	Central bearded dragon	Mesenchyme from dental tissues
Agamidae	SRR14455634	RNA-Seq	*Pogona vitticeps*	Central bearded dragon	Epithelium from dental tissues
Agamidae	SRR14455635	RNA-Seq	*Pogona vitticeps*	Central bearded dragon	Epithelium from dental tissues
Agamidae	SRR14455636	RNA-Seq	*Pogona vitticeps*	Central bearded dragon	Epithelium from dental tissues
Agamidae	SRR14455637	RNA-Seq	*Pogona vitticeps*	Central bearded dragon	Mesenchyme from dental tissues
Agamidae	SRR14455638	RNA-Seq	*Pogona vitticeps*	Central bearded dragon	Mesenchyme from dental tissues
Agamidae	SRR14455639	RNA-Seq	*Pogona vitticeps*	Central bearded dragon	Mesenchyme from dental tissues
Dactyloidae	GCF_000090745.1	Genome	*Anolis carolinensis*	Green anole	
Dactyloidae	SRR495265	RNA-Seq	*Anolis carolinensis*	Green anole	Adrenal Gland
Dactyloidae	SRR495266	RNA-Seq	*Anolis carolinensis*	Green anole	Adrenal Gland
Dactyloidae	SRR495267	RNA-Seq	*Anolis carolinensis*	Green anole	Adrenal Gland
Dactyloidae	SRR492481	RNA-Seq	*Anolis carolinensis*	Green anole	Adrenal Gland
Dactyloidae	SRR492482	RNA-Seq	*Anolis carolinensis*	Green anole	Adrenal Gland
Typhlopidae	GCA_022379055.1	Genome	*Anilios bituberculatus*	Prong-snouted blind snake	
Typhlopidae	SRR15431541	RNA-Seq	unclassified *Madatyphlops*		Mixture
Pythonidae	GCF_000186305.1	Genome	*Python bivittatus*	Burmese python	
Pythonidae	SRR11149661	RNA-Seq	*Python molurus*	Pythons	Spleen
Pythonidae	SRR11149662	RNA-Seq	*Python molurus*	Pythons	Skeletal muscle
Pythonidae	SRR11149663	RNA-Seq	*Python molurus*	Pythons	Brain
Pythonidae	SRR11149664	RNA-Seq	*Python molurus*	Pythons	Ovary
Pythonidae	SRR11149665	RNA-Seq	*Python molurus*	Pythons	Testis
Viperidae	GCA_003402635.1	Genome	*Protobothrops flavoviridis*	Habu	
Viperidae	DRR125541	RNA-Seq	*Protobothrops flavoviridis*	Habu	Fang forming tissue
Viperidae	DRR125542	RNA-Seq	*Protobothrops flavoviridis*	Habu	Venom gland
Viperidae	DRR125543	RNA-Seq	*Protobothrops flavoviridis*	Habu	Pit organ
Viperidae	DRR125544	RNA-Seq	*Protobothrops flavoviridis*	Habu	Nose
Viperidae	DRR125545	RNA-Seq	*Protobothrops flavoviridis*	Habu	Brain
Viperidae	DRR125546	RNA-Seq	*Protobothrops flavoviridis*	Habu	Eye
Viperidae	DRR125547	RNA-Seq	*Protobothrops flavoviridis*	Habu	Fetal fibroblast
Viperidae	DRR125548	RNA-Seq	*Protobothrops flavoviridis*	Habu	Venom gland
Viperidae	DRR125549	RNA-Seq	*Protobothrops flavoviridis*	Habu	Brain
Viperidae	DRR125550	RNA-Seq	*Protobothrops flavoviridis*	Habu	Spleen
Viperidae	DRR125551	RNA-Seq	*Protobothrops flavoviridis*	Habu	Lung
Viperidae	DRR125552	RNA-Seq	*Protobothrops flavoviridis*	Habu	Liver
Viperidae	DRR125553	RNA-Seq	*Protobothrops flavoviridis*	Habu	Kidney
Viperidae	DRR125554	RNA-Seq	*Protobothrops flavoviridis*	Habu	Pancreas
Viperidae	DRR125555	RNA-Seq	*Protobothrops flavoviridis*	Habu	Small Intestine
Viperidae	DRR125556	RNA-Seq	*Protobothrops flavoviridis*	Habu	Colon
Viperidae	DRR125557	RNA-Seq	*Protobothrops flavoviridis*	Habu	Stomach
Viperidae	DRR125558	RNA-Seq	*Protobothrops flavoviridis*	Habu	Heart
Viperidae	DRR125559	RNA-Seq	*Protobothrops flavoviridis*	Habu	Ovary
Viperidae	DRR125560	RNA-Seq	*Protobothrops flavoviridis*	Habu	Massseter muscle
Viperidae	GCA_018340635.1	Genome	*Bothrops jararaca*	Jararaca	
Viperidae	GCF_001527695.2	Genome	*Protobothrops*	Taiwanese habu	
Viperidae	GCA_003400415.2	Genome	*Crotalus viridis*	Prairie rattlesnake	
Viperidae	SRR7401978	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Liver

Viperidae	SRR7401979	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Liver
Viperidae	SRR7401980	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Kidney
Viperidae	SRR7401981	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Kidney
Viperidae	SRR7401982	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Liver
Viperidae	SRR7401983	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Liver
Viperidae	SRR7401984	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Kidney
Viperidae	SRR7401985	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Kidney
Viperidae	SRR7401986	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Pancreas
Viperidae	SRR7401987	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Tongue
Viperidae	SRR7401988	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Lung
Viperidae	SRR7401989	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Venom gland
Viperidae	SRR7401990	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Accessory venom gland
Viperidae	SRR7401991	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Testis
Viperidae	SRR7401992	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Pancreas
Viperidae	SRR7401993	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Shaker muscle
Viperidae	SRR7401994	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Stomach
Viperidae	SRR7401995	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Brain
Viperidae	SRR7401996	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Rictal gland
Viperidae	SRR7401997	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Ovary
Viperidae	SRR7401998	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Spleen
Viperidae	SRR7401999	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Blood
Viperidae	SRR7402000	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Liver
Viperidae	SRR7402001	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Kidney
Viperidae	SRR7402002	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Kidney
Viperidae	SRR7402003	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Kidney
Viperidae	SRR7402004	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Venom gland
Viperidae	SRR7402005	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Venom gland
Viperidae	SRR7402006	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Liver
Viperidae	SRR7402007	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Liver
Viperidae	SRR7402008	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Skin
Viperidae	SRR7402009	RNA-Seq	*Crotalus viridis*	Prairie rattlesnake	Pancreas
Viperidae	GCF_016545835.1	Genome	*Crotalus tigris*	Tiger rattlesnake	
Homalopsidae	GCA_017656035.1	Genome	*Myanophis thanlyinensis*		
Homalopsidae	SRR12802475	RNA-Seq	*Homalopsis buccata*	Masked water snake	Venom gland
Colubridae	GCF_001077635.1	Genome	*Thamnophis sirtalis*	Common garter snake	
Colubridae	SRR12915662	RNA-Seq	*Thamnophis conanti*		Duvernoy's gland
Colubridae	SRR1292619	RNA-Seq	*Boiga irregularis*		Duvernoy's gland
Colubridae	SRR14319402	RNA-Seq	*Tantilla nigriceps*	Plain black-headed snake	Duvernoy's gland
Colubridae	SRR14319403	RNA-Seq	*Tantilla nigriceps*	Plain black-headed snake	Duvernoy's gland
Colubridae	SRR14319404	RNA-Seq	*Tantilla nigriceps*	Plain black-headed snake	Duvernoy's gland
Colubridae	SRR14319405	RNA-Seq	*Tantilla nigriceps*	Plain black-headed snake	Duvernoy's gland
Colubridae	SRR14319406	RNA-Seq	*Tantilla nigriceps*	Plain black-headed snake	Duvernoy's gland
Colubridae	SRR14319407	RNA-Seq	*Tantilla nigriceps*	Plain black-headed snake	Duvernoy's gland
Colubridae	GCF_001185365.1	Genome	*Pantherophis guttatus*	Corn snake	
Colubridae	SRR9596700	RNA-Seq	*Pantherophis guttatus*	Corn snake	Brain
Colubridae	SRR9596701	RNA-Seq	*Pantherophis guttatus*	Corn snake	Brain
Colubridae	SRR9596706	RNA-Seq	*Pantherophis guttatus*	Corn snake	Heart
Colubridae	SRR9596707	RNA-Seq	*Pantherophis guttatus*	Corn snake	Liver
Colubridae	SRR9596708	RNA-Seq	*Pantherophis guttatus*	Corn snake	Dorsal skin
Colubridae	SRR9596709	RNA-Seq	*Pantherophis guttatus*	Corn snake	Ventral skin
Colubridae	SRR9596710	RNA-Seq	*Pantherophis guttatus*	Corn snake	Vomeronasal Organ
Colubridae	SRR9596711	RNA-Seq	*Pantherophis guttatus*	Corn snake	Vomeronasal Organ
Colubridae	SRR9596712	RNA-Seq	*Pantherophis guttatus*	Corn snake	Cerebellum
Colubridae	SRR9596713	RNA-Seq	*Pantherophis guttatus*	Corn snake	Testis
Colubridae	SRR9596714	RNA-Seq	*Pantherophis guttatus*	Corn snake	Ovary
Colubridae	SRR9596715	RNA-Seq	*Pantherophis guttatus*	Corn snake	Liver
Colubridae	SRR9596716	RNA-Seq	*Pantherophis guttatus*	Corn snake	Kidney
Colubridae	SRR9596717	RNA-Seq	*Pantherophis guttatus*	Corn snake	Cerebellum
Colubridae	GCA_019677565.1	Genome	*Pituophis catenifer*	Copher snake	
Colubridae	GCA_012654085.1	Genome	*Pantherophis obsoletus*	Western rat snake	
Colubridae	GCF_009769535.1	Genome	*Thamnophis elegans*	Western terrestrial garter snake	
Colubridae	SRR12802480	RNA-Seq	*Rhabdophis subminiatus*	Red-necked keelback	Duvernoy's gland
Dipsadidae	GCA_003457575.1	Genome	*Thermophis baileyi* (Closed species to *C. lineatus*)		Duvernoy's gland
Dipsadidae	SRR14319401	RNA-Seq	*Conophis lineatus*		Duvernoy's gland
Dipsadidae	SRR12802479	RNA-Seq	*Heterodon nasicus*	Western hognose snake	Duvernoy's gland

Dipsadidae	SRR12802481	RNA-Seq	*Helicops leopardinus*		Duvernoy's gland
Dipsadidae	SRR1292610	RNA-Seq	*Hypsiglena*		Duvernoy's gland
Lamprophiida	SRR12802476	RNA-Seq	*Psammophis subtaeniatus*	Western yellow-bellied sand	Venom gland
Lamprophiida	SRR12802477	RNA-Seq	*Psammophis schokari*	Schokari sand racer	Venom gland
Lamprophiida	SRR12802478	RNA-Seq	*Malpolon monspessulanus*	Montpellier snake	Venom gland
Elapidae	GCA_000516915.1	Genome	*Ophiophagus hannah*	King cobra	
Elapidae	GCA_015471245.1	Genome	*Laticauda colubrina*	Yellow-lipped sea krait	
Elapidae	GCF_900518725.1	Genome	*Notechis scutatus*	Mainland tiger snake	
Elapidae	SRR8206942	RNA-Seq	*Naja haje*	Egyptian cobra	Venom gland
Elapidae	GCA_009733165.1	Genome	*Naja naja*	Indian cobra	
Elapidae	SRR8206943	RNA-Seq	*Naja naja*	Indian cobra	Venom gland
Elapidae	SRR8206944	RNA-Seq	*Naja kaouthia*	Monocled cobra	Venom gland
Elapidae	SRR11050878	RNA-Seq	*Naja siamensis*	Thai spitting cobra	Venom gland
Elapidae	SRR11050880	RNA-Seq	*Naja nubiae*	Nubian spitting cobra	Venom gland
Elapidae	SRR11050881	RNA-Seq	*Naja nivea*	Cape cobra	Venom gland
Elapidae	This study	Genome	*Naja atra*	Chinese cobra	Muscle
Elapidae	SRR11050884	RNA-Seq	*Naja atra*	Chinese cobra	Venom gland
Elapidae	SRR11050887	RNA-Seq	*Naja sumatrana*	Equatorial spitting cobra	Venom gland
Elapidae	SRR11050889	RNA-Seq	*Hemachatus haemachatus*	Rinkhals	Venom gland
Elapidae	SRR11050890	RNA-Seq	*Naja philippinensis*	Philippine spitting cobra	Venom gland
Elapidae	SRR11050879	RNA-Seq	*Naja pallida*	Red spitting cobra	Venom gland
Elapidae	SRR11050882	RNA-Seq	*Naja nigricollis*	Black-necked spitting cobra	Venom gland
Elapidae	SRR11050883	RNA-Seq	*Naja mossambica*	Mozambique spitting cobra	Venom gland
Elapidae	SRR11050885	RNA-Seq	*Naja annulifera*	Snouted cobra	Venom gland
Elapidae	SRR11050886	RNA-Seq	*Walterinnesia aegyptia*	Desert cobra	Venom gland
Elapidae	SRR11050888	RNA-Seq	*Naja subfulva*	Brown forest cobra	Venom gland
Hydrophiidae	GCA_019472885.1	Genome	*Hydrophis curtus*	Shaw's sea snake	
Hydrophiidae	SRR11659669	RNA-Seq	*Hydrophis curtus*	Shaw's sea snake	Venom gland
Hydrophiidae	SRR11659670	RNA-Seq	*Hydrophis curtus*	Shaw's sea snake	Venom gland
Hydrophiidae	SRR11659671	RNA-Seq	*Hydrophis curtus*	Shaw's sea snake	Venom gland
Hydrophiidae	GCA_019473425.1	Genome	*Hydrophis cyanocinctus*	Blue-banded sea snake	
Hydrophiidae	SRR11659657	RNA-Seq	*Hydrophis cyanocinctus*	Blue-banded sea snake	Venom gland
Hydrophiidae	SRR11659658	RNA-Seq	*Hydrophis cyanocinctus*	Blue-banded sea snake	Venom gland
Hydrophiidae	SRR11659659	RNA-Seq	*Hydrophis cyanocinctus*	Blue-banded sea snake	Venom gland

Table S2 Crystallographic and refinement statistics

	***Naja atra* svPDE** apo form (unliganded) PDB accession: 5GZ4	***Naja atra* svPDE** AMP-complexed form (liganded) PDB accession: 5GZ5
Data collection		
Wavelength (Å)	0.97	1.00
Space group	P21212	P21212
Cell dimensions		
a, *b*, *c* (Å)	171.679, 65.876, 89.615	171.207, 65.612, 88.675
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	30 - 2.55 (2.64 – 2.55)*	30 - 2.09 (2.18 – 2.09)
R_{merge}	0.111 (0.560)	0.059 (0.570)
$CC_{1/2}$	0.92 (0.723)	0.945 (0.785)
$I/\sigma I$	13.1 (2.0)	22.5 (2.5)
Completeness (%)	99.8 (99.8)	99.1 (98.9)
Redundancy	4.5 (4.7)	4.8 (4.9)
Refinement		
Resolution (Å)	29.43 - 2.55 (2.64 – 2.55)	27.88 - 2.09 (2.16 – 2.09)
No. of reflections	33351	58940
R_{work}/ R_{free}	0.241/0.289	0.181/0.217
No. atoms		
Protein	6450	6255
Ligand/ion	162	157
Water	677	804
B-factors		
Protein	49.39	37.91
Ligand/ion	52.42	56.27
Water	40.52	42.23
R.m.s deviations		
Bond lengths (Å)	0.010	0.012
Bond angles (°)	1.78	1.52
Validation (MolProbity[1])		
Ramachandran favored (%)	94.00	96.00
Ramachandran outliers (%)	0.88	0.26

* Highest resolution shell is shown in parenthesis.

1. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.

Table S3 Direct biochemical measurement of the substrate specificity

	Protein	**svPDE***	**svPDE[2]**	**ENPP3[3]**	**ENPP1[4]**
	Species	*Naja atra*	*Trimeresurus stejneger*	*Homo sapiens*	*Homo sapiens*
Reaction Condition		37°C, pH8	37°C, pH7.4	37°C, pH7.5	NA
ATP	K_m (μM)	336.0	360 ± 0.01	61.50 ± 6.40	144.5 ± 36.0
	k_{cat} (s^{-1})	44.0	2.68 ± 0.13	1.52 ± 0.41	7.8 ± 0.8
ADP	K_m (μM)	654.0	91 ± 0.005	NA	NA
	k_{cat} (s^{-1})	37.0	0.23 ± 0.01	NA	NA
NAD	K_m (μM)	539.0	240 ± 0.01	NA	NA
	k_{cat} (s^{-1})	608.0	0.82 ± 0.03	NA	NA
GTP	K_m (μM)	2160.0	NA	123.70 ± 6.60	NA
	k_{cat} (s^{-1})	717.9	NA	3.37 ± 0.31	NA
CTP	K_m (μM)	5750.0	NA	120.30 ± 10.50	NA
	k_{cat} (s^{-1})	1502.0	NA	6.36 ± 0.30	NA
UTP	K_m (μM)	2730.0	NA	120.20 ± 26.80	NA
	k_{cat} (s^{-1})	684.6	NA	9.14 ± 0.68	NA

* This study.

2. Peng, L., et al., *Purification and partial characterization of a novel phosphodiesterase from the venom of Trimeresurus stejnegeri: inhibition of platelet aggregation.* Biochimie, 2011. **93**(9): p. 1601-9.
3. Gorelik, A., et al., *Structural basis for nucleotide recognition by the ectoenzyme CD203c.* FEBS J, 2018. **285**(13): p. 2481-2494.
4. Albright, R.A., et al., *Molecular basis of purinergic signal metabolism by ectonucleotide pyrophosphatase/phosphodiesterases 4 and 1 and implications in stroke.* J Biol Chem, 2014. **289**(6): p. 3294-306.

Table S4 Statistics of PacBio CCS reads

Library ID	Total pass bases (bp)	Total pass reads (#)	Mean length of pass reads (bp)	Max. length of pass reads (bp)	N50 length of pass reads (bp)	% reads >10Kb	% reads >20Kb
m64144_201215_090635	29,676,086,054	2,375,903	12,490	39,549	12,666	87.91	0.50
m64144_201216_153226	26,959,253,579	2,187,719	12,322	36,670	12,481	86.49	0.45
Total	56,635,339,633	4,563,622	12,410	39,549	12,578	87.23	0.48

Table S5 Statistics of Illumina paired-end reads

Total reads	Clean reads	Total bases (bp)	Clean bases (bp)	Q20 (%)	Q30 (%)	GC (%)
568,048,858	559,566,268	85,207,328,700	78,096,046,566	97.37	92.63	41.10

Table S6 Statistics of draft assembly

Statistics	De novo assembly of PacBio reads		Assembly polished with Illumina reads	
	Contig Length(bp)	Contig Number	Contig Length(bp)	Contig Number
N50	29,710,036	18	29,709,188	18
N60	20,075,995	26	20,074,295	26
N70	10,560,158	38	10,559,546	38
N80	4,204,041	65	4,203,549	65
N90	1,530,490	140	1,530,247	140
Longest	111,070,803	1	111,070,038	1
Total	1,879,633,414	847	1,879,550,836	847