

Citation: Chapeaurouge A, Reza MA, Mackessy SP, Carvalho PC, Valente RH, Teixeira-Ferreira A, et al. (2015) Interrogating the Venom of the Viperid Snake *Sistrurus catenatus edwardsii* by a Combined Approach of Electrospray and MALDI Mass Spectrometry. PLoS ONE 10(5): e0092091. doi:10.1371/journal.pone.0092091

Academic Editor: Partha Mukhopadhyay, National Institutes of Health, UNITED STATES

Received: December 5, 2013

Accepted: February 17, 2014

Published: May 8, 2015

Copyright: © 2015 Chapeaurouge et al. This is an open access article distributed under the terms of the <u>Creative Commons Attribution License</u>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by Biomedical Research Council, Agency for Science, Technology and Research, Singapore and grants from Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq to J. Perales), and Programa de Desenvolvimento Tecnológico em Insumos para Saúde (PDTIS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. **RESEARCH ARTICLE**

Interrogating the Venom of the Viperid Snake *Sistrurus catenatus edwardsii* by a Combined Approach of Electrospray and MALDI Mass Spectrometry

Alex Chapeaurouge^{1,2}, Md Abu Reza^{2#}, Stephen P. Mackessy³, Paulo C. Carvalho^{1,4}, Richard H. Valente¹, André Teixeira-Ferreira¹, Jonas Perales¹, Qingsong Lin², R. Manjunatha Kini^{2,5}*

Laboratório de Toxinologia, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, RJ, 21045–900, Brazil,
Department of Biological Sciences, 14 Science Drive 4, National University of Singapore, Singapore,
Singapore, 3 School of Biological Sciences, University of Northern Colorado, 501 20th St., CB 92,
Greeley, Colorado, 80639–0017, United States of America, 4 Laboratory for Proteomics and Protein
Engineering, Carlos Chagas Institute, Fiocruz, Curitiba, PR, 81350–010, Brazil, 5 Department of
Biochemistry and Molecular Biology, Medical College of Virginia, Virginia Commonwealth University,
Richmond, Virginia, 23298–0614, United States of America

Current address: University of Rajshahi, Department of Genetic Engineering and Biotechnology, Rajshahi,
6205, Bangladesh

* dbskinim@nus.edu.sg

Abstract

The complete sequence characterization of snake venom proteins by mass spectrometry is rather challenging due to the presence of multiple isoforms from different protein families. In the present study, we investigated the tryptic digest of the venom of the viperid snake *Sistrurus catenatus edwardsii* by a combined approach of liquid chromatography coupled to either electrospray (online) or MALDI (offline) mass spectrometry. These different ionization techniques proved to be complementary allowing the identification a great variety of isoforms of diverse snake venom protein families, as evidenced by the detection of the corresponding unique peptides. For example, ten out of eleven predicted isoforms of serine proteinases of the venom of *S. c. edwardsii* were distinguished using this approach. Moreover, snake venom protein families not encountered in a previous transcriptome study of the venom gland of this snake were identified. In essence, our results support the notion that complementary ionization techniques of mass spectrometry allow for the detection of even subtle sequence differences of snake venom proteins, which is fundamental for future structure-function relationship and possible drug design studies.

Introduction

Snake venoms not only represent rich sources of biologically active peptides and proteins but also serve as versatile platforms for the discovery and development of drug lead substances [1].

Competing Interests: The authors have declared that no competing interests exist.

PLOS ONE

Significant progress in the investigations of snake venoms has recently been witnessed by different proteomics studies in this field. The combined transcriptome and proteome analysis of the venom of *Cerberus rynchops*, for example, revealed a very low complexity venom composition and a novel snake venom protein family called veficolins; function of veficolins has been hypothesized to be related to the inhibition of platelet aggregation [2]. Likewise, investigations into the venom of the ocellated carpet viper *Echis ocellatus* pointed to a pronounced role of transcriptional and posttranslational mechanisms on determining the final venom composition, as evidenced by a significant divergence between predicted toxin clusters found in the transcriptome and peptide sequences identified in the corresponding venom proteome [3]. A comparative proteome analysis of the venoms of terrestrial Toxicocalamus longissimus and a closely related marine species Hydrophis cyanocinctus indicates a pronounced reduction of the molecular diversity of the venom components of the marine snake as compared to the venom proteome of its terrestrial relative [4]. The authors reason that molecular economy of the toxin arsenal has been implemented as an evolutionary response to selective pressures from different environmental challenges. To predict possible structure function relationships of the various proteins of the corresponding venom, a complete picture of the sequences of the different protein families and their isoforms is of major importance. Extensive sequence coverage of the venom proteome can be accomplished using a combined approach of electrospray and MALDI ionization mass spectrometry. In the present study, we have used this approach to characterize the venom proteome of the pitviper Sistrurus catenatus edwardsii (Desert Massasauga Rattlesnake), a subspecies of Sistrurus catenatus, which is primarily encountered in dry and desert grasslands of the southwestern North American prairies [5, 6]. A comparative study of the venom proteomes of four different Sistrurus taxa has revealed an overview of the different protein families of the corresponding venoms, as evidenced by BLAST analysis of the detected sequences [7]. The transcriptome of the venom gland of S. c. edwardsii has also been characterized and serves as an exhaustive source for protein sequence investigations of the venom proteome [8]. Based on the identification of unique peptides of the corresponding proteins we were able to distinguish ten out of eleven predicted isoforms of serine proteinases and all five predicted metalloproteinase isoforms, together with a disintegrin. We also encountered the snake venom protein families C-type lectin, cysteine rich secretory protein, nerve growth factor, phospholipase A2, bradykinin-potentiating protein, and L-amino acid oxidase, previously described in the transcriptome of S. c. edwardsii. In addition, our analysis revealed the presence of snake venom protein families not detected in the venom gland transcriptome or previous studies, including glutaminyl cyclase, renin-like aspartic protease, and ecto-5'-nucleotidase. These results support the view that an in-depth analysis of the venom proteome is complementary to transcriptomic venom gland studies and will improve our understanding of the interplay of the different venom proteins on the target prey.

Materials and Methods

Venom extraction and Ethics statement

Specimens of *Sistrurus catenatus edwardsii* (Desert Massasauga) were collected in Lincoln County, Colorado, USA under permits granted by the Colorado Division of Wildlife to Stephen P. Mackessy (permits #0456, 06HP456). Venom was extracted manually [9] from 4 adult snakes from the same metapopulation in southeastern Colorado; venoms were pooled, centrifuged and lyophilized. Snakes were then PIT-tagged, returned to the exact locality and released. All procedures were permitted by the University of Northern Colorado Institutional Animal Care and Use Committee as detailed in UNC IACUC protocol #0702. No animals were sacrificed and no suffering of animals occurred during this study.

Tryptic digestion of the venom

Lyophilized crude venom (600 μ g) was initially dissolved in 600 μ l of ammonium bicarbonate (50 mM) and precipitated with three volumes of ice-cold acetone for 3 h at -20°C. In the following, the sample was spun down at 14000 rpm for 10 min and the pellet brought up in 560 μ l of ammonium bicarbonate (50 mM). Afterwards, 12 μ l of 1% ProteaseMAX (in 50 mM ammonium bicarbonate) surfactant was added to the sample solution followed by reduction with 8 μ l of 0.5 M DTT (at 56°C for 20 min) and alkylation with 16 μ l of 0.55 M iodoacetamide at room temperature for 30 min in the dark. Finally, the venom was subjected to digestion (at 37°C for 12 h) by adding 10 μ l of trypsin (1 μ g/ μ l in 50 mM acetic acid) and 6 μ l of 1% ProteaseMAX surfactant to enhance the enzymatic performance of trypsin.

Chromatography and mass spectrometry

The tryptic digest was separated on a C-18 reversed phase column (Agilent Zorbax 300SB-C18 1.0 x 150 mm x 3.5 µm) by running a linear gradient from 0% acetonitrile to 72% acetonitrile in 120 min applying a flow rate of 40 μ /min (solvent A contains H₂O / 0.1% TFA, solvent B contains 80% ACN / 0.1% TFA) on a LC Packings Ultimate HPLC system (Dionex, Vernon Hills, IL). During the chromatographic run approximately 130 fractions were manually collected in Eppendorf vials. After reducing the volume in each vial to approximately 1 µl on a Speedvac (Savant SC 110A), samples were spotted onto the MALDI sample plate. Approximately 0.3 µl of the sample solution was mixed with the same volume of a saturated matrix solution (α -cyano-4-hydroxycinnamic acid, (Aldrich, Milwaukee, WI) 10 mg/ml in 50% acetonitrile/0.1% trifluoroacetic acid) on the target plate and allowed to dry at room temperature (dried-droplet method). Raw data for protein identification were obtained on a AB Sciex 5800 (AB Sciex, Foster City, CA) and a 4700 Proteomics Analyzer (Applied Biosystems, Foster City, CA). Typically, 1600 shots were accumulated for spectra in MS mode while 3500 shots were accumulated for spectra in MS/ MS mode. Up to twenty of the most intense ion signals with a signal to noise ratio above 2 were selected as precursors for MS/MS acquisition excluding common trypsin autolysis and keratin peaks. External calibration in MS mode was performed using a mixture of six singly charged peptides: des-Arg1-Bradykinin (m/z = 904.468), angiotensin I (m/z = 1296.685), Glu1-fibrinopeptide B (m/z = 1570.677), ACTH (1–17 clip) (m/z = 2093.087), ACTH (18–39 clip) (m/z = 2465.199), and ACTH (7-38 clip) (m/z = 3,657.929). MS/MS spectra were externally calibrated using known fragment ion masses observed in the tandem mass spectrum of Glu1-fibrinopeptide B.

Tryptic peptides were also separated on an Easy nLC II (Thermo Scientific) nanoflow HPLC system connected to an LTQ-Orbitrap XL mass spectrometer (Thermo, Bremen, Germany) equipped with a nanoelectrospray ion source. Peptides were initially loaded onto a trap column (100 μ m x 2 cm) packed in-house with C₁₈ resin (5 μ m, 100 Å pore, Magic C₁₈ AQ, Bruker-Michrom, Auburn, CA) and separated on an RP HPLC column (C_{18} , 75 µm x 30 cm) using a linear gradient from 98% solvent A (H₂O, 0.1% formic acid) to 60% solvent B (ACN, 0.1% formic acid) over 162 min. Precursor scans were performed in the Orbitrap mass detector at a resolution of 60,000 in the mass range of 300 m/z to 1700 m/z, while MS/MS scans were acquired in the linear trap ("high-low"). With an exclusion of singly charged ions, up to ten of the most intense precursor ions were subjected to product ion scans using CID with a normalized collision energy of 35%. Moreover, MS/MS scans were only triggered for precursor ions having a minimum signal threshold of 10,000 counts. Precursors that were selected for MS/ MS scans were dynamically excluded for 30 sec from a repeated product ion scan within a ± 10 ppm mass error. Different HPLC separations were performed where the fragmentation of precursor ions was induced using CID only as well as an approach of alternating CID and ETD, in which successively the same precursor ion was fragmented by CID and ETD.

Data analysis

Database searches of the mass spectra acquired on the MALDI mass spectrometers were searched against all entries of NCBInr (<u>www.ncbi.nlm.nih.gov/index.html</u>) using the Mascot software (<u>www.matrixscience.com</u>) and against an in-house created snake venom database using Mascot (Mascot, version 2.1). The following search parameters were used: No restrictions on species of origin or protein molecular weight, semi-tryptic cleavage products, two tryptic missed cleavages allowed, variable modifications of cysteine (carbamidomethylation) and methionine (oxidation), and pyroglutamate formation at N-terminal glutamine of peptides.

Electrospray data were analyzed using the Peaks (Peaks Studio 5.3) and ProLuCID [10] search engines, respectively, against an in-house created snake venom database (2580 entries). The search parameters of the Peaks search engine were sulfation (serine, threonine), phosphorylation (serine, threonine) and tyrosine), deamidation (glutamine, asparagine), dehydration (serine, threonine) oxidation (methionine, tryptophan, and histidine) and acetylation at N-terminal of peptides, with a maximum number of 3 modifications per peptide allowed. Search parameters of ProLuCID were fixed modification of cysteine (carbamidomethylation), variable modifications of methionine (oxidation), a precursor tolerance of 50 ppm and allowance for semi-tryptic identifications. Peptide spectrum matches obtained by ProLuCID were then validated by the Search Engine Processor with the default parameters previously described [11]. All identified peptides were further manually verified.

Results and Discussion

Envenomation by viperid snakes frequently manifests as a complex medical syndrome dominated by hemorrhagic and inflammatory processes triggered by the combined enzymatic actions of metalloproteinases, serine proteinases and phospholipases $A_2 \left[\frac{12}{12} - \frac{14}{14} \right]$, as well as by the detrimental effects of C-type lectins (CLP) on platelet function [15]. The snake venom metalloproteinases (SVMP) are classified in four different major groups (PI, PII, PIII, and PIV) based on their final domain composition after posttranslational modification of the corresponding multidomain precursor protein [16, 17]. Functionally, a broad spectrum of biological activities have been attributed to SVMPs, including hemorrhagic, inflammatory and myonecrotic effects [14, 18]. To date, only a few studies have noted the presence of peptides related to the prodomain of SVMPs in the venom [19, 20] and it might be that in most cases the prodomain of the precursor protein is enzymatically removed before its secretion into the venom gland. Only a single (identical) peptide of the prodomain of SVMP isoforms 1 and 3 (Fig 1) was identified, suggesting that the prodomain is proteolytically removed in S c. edwardsii before being exocytosed to the venom gland lumen. The metallo-, disintegrin-, and cysteine rich domains of the four isoforms that belong to the PIII class of metalloproteinases revealed evenly distributed sequence coverage (Fig 1), supporting the view that these domains are efficiently translated. Similarly, we were able to identify tryptic fragments of the predicted [8] PII (isoform 6) and disintegrin of the venom proteome, with sequence coverages of 36% and 71%, respectively (Fig 1). The presence of proteotypic peptides of the corresponding isoforms clearly revealed the existence of these different proteins in the venom proteome (Fig 2). Of further particular note is the presence of protein sequences that went undetected in the transcriptome analysis but could be identified in the venom proteome. These sequences belong to 26 different SVMPs identified as PII and PIIIs and disintegrins of snakes phylogenetically closely related to S. c. edwardsii (Table 1). Interestingly, during a recent investigation of the transcriptome and proteome of the cryptic snake Drysdalia coronoides, the SVMPs of the venom were initially only identified in the proteome and only the implementation of gene-specific 3'RACE primers of the corresponding signal peptides of the targeted proteins revealed the cDNA sequence [21]. These findings might point to a general difficulty to characterize



Protein (Coverage %)	Sequence	
Metallo- proteinase 1 (55.0)	propeptide metalloprotease-domain active site disintegrin-domain cysteine rich-domain 11 132 179 HEHH 373 392 467	587 591
Metallo- proteinase 2 (48.0)	propeptide metalloprotease-domain active site disintegrin-domain cysteine rich-domain	⁵⁸⁸ 592
Metallo- proteinase 3 (52.4)	propeptide metalloprotease-domain disintegrin-domain cysteine rich-domain	578 582
Metallo- proteinase 7 (56.2)	propeptide metallo protease-domain active site disintegrin-domain cysteine rich-domain	499 503
Metallo- proteinase 6 (36.0)	propeptide metalloprotease-domain _{active site} disintegrin-doma	ain 444
Disintegrin (71.5)	disintegrin-domain	176
Phospho- lypase A 2 (84.2)	primary metal binding site (Ca ²⁺) phospholypase A2 - domain Y G G D 115	122
Bradykinin-	atrial natriuretic pepti	tide

Fig 1. Sequence coverages of some of the predicted venom gland proteins of *S. c. edwardsii* as revealed by the combined approach of MALDI and ESI tandem mass spectrometry. Protein sequences including specific domains are indicted by colored bars; below these, corresponding peptides identified by ESI (black lines) and MALDI (red lines) are indicated.

doi:10.1371/journal.pone.0092091.g001

6 2 1 7		60
5 3	MIEVLLVTICLAVFPYQGSSIILESGNVNDYEVVYPRKVTALPKGAVQPNYEDAMQYEFK	60
6 2 1 7	VNGEPVVLHLEKNKGLFSEDYSETHYSPDGREITTYPLVEDHCHYLGRIQNDADSTASLS VNGEPVVLHLEKNKELFSEDYSETHYSPDGRKITTYPSDEDHCYYHGRIQNDADSTASIS MNGEPVVLHLEKNKGLFSKDYSETHYSPDGRQITTYPMIEDHCYYHGCIQNDADSTASIS RIENDADSTASIS	120 120
5 3		120
6 2 1 7 5	ACNGLKGYFK-LQGEMYLIEPLKLPDNEAHAVYKYENVEKEDEAPKMCGVTQNWESDKLI ACSGLKRTFQAFGGEMYLIEPLKLPDSEAHAVFKYENVEKEDEAPKMCGVTQNWESYEPI ACNGLKGHFK-LQGEMYLIEPLKLPDSEAHAVYKYENIEKEDEAPKMCGVTQNWESYEPI ACNGLQGHFK-LQGEMYLIEPLKLPDSEAHAVFKYENVEKEDEAPKMCGVTQNWESYEPI	180 179
3	ACNGLKGHFK-LQGEMYLIEPLELPDSEAHAVFKYENVEKEDEAPKMCGVTQTNWKSYEP	179
6 2 1 7 5	KKASYLDVTAEQQRFPQRHIELVVVADHRMFTKYNSNLNTVRIWVHEIVNIINVFYRS KKASQLVVTAEQQRHLNNFRYIELVIVADYRMFTKYNRKLTEVKTWVYEIVNTLNEIYRY KKASQLNLTPEQQAYLDAKKYVEFVVVLDHGMYTKYKDDLDKIKTRIYEIVNTMNEIYIP KKASQSNLTPEQQRYLNAKKYVKLFLVADYVMYLKYGRNLTAVRTRMYDTVNVITPIYHR	240 239
3	IKKPSHLVVTAEQQRLTKR YVELVIVADHGM FMKHNETLR TWVFQMVNTINEIYIP	235
6 2 1 7 5	LNIHVSLTDLEIWSNQDQINVQSAAADTLEAFGDWRETVLLNRISHDNAQLLTAIELDEE LYIRVALVGLEIWSNGDLSNVTLSADDTLDSFGEWRKRDLLRRKSHGNAQLLTAIDFNGT LNIRVALVRLEIWSNGDLINVSSAANVTLDSFRNWRVTDLLRRKSHDNAQLLTTDFDGD MNIHVALVGLEIWSNRDKIIVQSSPDVTLDLFAKWRATDLLRRKSHDNAQLLTGINFNGP	300 299 192
3	LNIRFALVGLEIWSNGDKITVQSSAHNTLHLFGNWRKRVLLKRKKHDNAQLLTSIDFDGR	
6 2 1 7 5	TIGLAHVGSMCDPKLSTGIVQDHSAINLWVAVTMAHELGHNLGMDHDGNQCYCDTDSCIM TIGLAHVASMCELKR STGIVQDHSPINLLVAVTMAHEMGHNLGIR HDKKYCTCGGYSCIM TVGLAYMR SMCQPR GSVGVIQEHSTINLL MAVTMAH EMGHNLGMSHDGNQCHCGAPSCIM TAGLAYLGGICNTMYSAGIVQDHSKIHHLVAIAMAHEMGHNLGMDHDKETCTCGGTRPCVM	360 359 252
3	TVGLAHVSSMCNQKLSTGVIQDHSAINLVMAVTMAHELGHNLGMDHDIKYCTCGAPACVM	355
6 2 1 7 5 3	SEELSEQLSFEFSDCSLNEYQTYLTDHNPQCMLNEPLRTDTVSGNELLEAG EECDC SAVLSHQASKYFSNCSYNQYWNYINF YKPQCILNEPLR TDIVSPPVCGNELLEAGEECDC ADRLSHQPSTQFSDCSEKYCRTYLINSRPQCILNEPLLTDIVSPPVCGNELLEEGEECDC AGTLSCEASFLFSDCSQKEHREFLIKNNPQCILKKFLKTDVVSPAVCGNYFVEVGEECDC GNTLSEQPSYGFSNCSQNQYQTFIINYNPQCILNEPLPTDIVSPPVCGNELLEVGEECDC	420 419 312
6 2 1 7 5 3	GSPANPCCDAATCKLRPGAQCAEGLCCDQCRFIKKGKICRRARGD-NPDDRCTGQSA GSPATCQYQCCNATTCKLHSWVECESGECCEQCIFKGAGTECRAASSECDLPENCTGHSG GSPANCQNPCCDAATCKLTPGSQCAKGLCCDQCRFKGAGTECRAAKDDCDMAPVCTGRSA GSPRTCRDPCCDAATCKLRQGAQCAEGLCCDQCRFKGAGTECRAAKDDCDMADVCTGRSA GKPGTQCEDGECCDRCQLRRAGTECRPARSECDIAESCTGQSA GSPNTCRDPCCDATTCKQHSWVECESGECCEQCKFKGAGTECRAARSECDIAESCTGQSA	480 479 372 43
6 2 1 7 5 3	DCPRNRFHA	540 539 431 103
6 2 1 7 5 3	CRKENGRKIPCAPEDVKCGRLYCSYNFNGNQIPCVPYYTRRDENMGMVEPGTKCGDGKVC CRKENGTKIPCEPEDEKCGRLFCFPNKPGKKNICNVIYTPTDEDIGMVLPGTKCGRGKVC CRKEQNTKIACEPQDVKCGRLYCFPNSPETKNPCNIYYSPNDEDKGMVLPGTKCADGKAC CRRENGRIFPCAPQDVKCGRLFCVMGPTGHTISCQSTYSQNDLDIGMVDLGTKCGDGRVC CRREKARIFPCAPQDVKCGRLFCEFGHYSCQYQYS-GDPDFGMVDHGTKCADGKVC	599 491 163
6 2 1 7 5 3	S-NGQCVDVNIAY 612 S-NGGCVDVTATY 611 S-NGQCVDVTPY 503 NSNRQCVDVNTAY 176 K-NRQCVDVTTAY 602	

Fig 2. Sequence alignment of the metalloproteinases of the venom of S. c. edwardsii. Different colors indicate the unique peptides identified by tandem mass spectrometry of the corresponding proteins.

doi:10.1371/journal.pone.0092091.g002

the relatively large SVMP sequences fully in the transcriptome of the venom gland without the use of amplification techniques. However, it is worth mentioning that a recent in-depth transcriptome analysis of the venom of the eastern diamond rattlesnake *Crotalus adamanteus* produced full-lengths SVMP sequences by using next-generation sequencing including the Illumina technology [22] The high abundance of metalloproteinases in the venom of *S. c. edwardsii* is in line with a previous transcriptome analysis [8] and point to an explicit role of accelerated evolution on the development of distinct metalloproteinase isoforms [23].

Snake venom serine proteinases (SVSPs) primarily affect the hemostatic system of prey organisms and often show fibrinogenolytic and fibrinolytic activities. Since this is similar to the action of thrombin on fibrinogen, SVSPs have also been known as thrombin-like enzymes (TLEs) [24, 25]. In addition, SVSPs also act on kininogen and platelet receptors [26]. While most SVSPs exist as monomers, dimeric forms have been detected in the venom of the viper A. b. brevicaudus [27]. One of these is brevinase, a heterodimeric enzyme with a covalent disulfide link between the two monomers. The analysis of the transcriptome of the venom gland of S. c. edwardsii predicted the presence of eleven SVSP-specific isoforms. Previous studies on the protein coding region of SVSPs in pitvipers have noted a trend towards accelerated evolution of this protein family, a result which is also found in the venom of S. c. edwardsii, as evidenced by the ratio (0.99) between non-synonymous and synonymous substitutions of the exon sequences of the mRNA transcripts [8]. Such sequence variations are likely related to different pharmacological activities. The combined electrospray and MALDI ionization analysis of the proteome resulted in the identification of ten out of eleven predicted SVSP isoforms, as evidenced by the detection of unique peptides (Fig 3) with a sequence coverage between 35% (isoform 8) and 82% (isoform 1). Again, we observed the presence of additional peptides that match sequences of SVSPs from related snake venoms (e.g. C. adamanteus) and that were not identified in the transcriptome analysis of S. c. edwardsii (Table 1). Taken together, these additional ten identified isoforms raises the total number of isoforms of SVSPs to a total number to twenty, reinforcing the idea that this protein family has evolved in an accelerated manner, producing an elevated number of isoforms.

Phospholipase A_2 (PLA₂s) are functionally characterized by their multiple pharmacological activities such as cardiotoxic, neurotoxic, myotoxic, antiplatelet and anticoagulant effects [28, 29]. They enzymatically cleave the second ester bond of the glycerol ester and represent one of the most extensively studied snake venom families. The venom proteome revealed the presence of only one PLA₂ protein (Fig 1), consistent with the transcriptome analysis of the venom gland [8]. However, an additional single peptide that showed sequence identity to a PLA₂ from *S. c. tergeminus* (Table 1) was also found. It appears that, contrary to other species, where PLA₂s are present in multiple isoforms, the venom of *S. c. edwardsii* was not under evolutionary pressures selecting for the evolution of a pronounced diversity of PLA₂s.

The cysteine-rich-secretory proteins (CRISP) are widely distributed in snake venoms, particularly in Viperidae and Elapidae. The biological activity of some is related to the inhibition of the cyclic nucleotide-gated ion channels as well as L-type Ca²⁺ and BK_{Ca} K⁺ channels [30]. For example, triflin and ablomin (from the pitviper *Gloydius blomhoffii*) block L-type Ca²⁺ channels that lead to contraction of smooth muscle [31, 32]. However, scientists are only beginning to understand the full scope of biological and pharmacological effects of this protein family. We identified the predicted CRISP protein in the venom of *S. c. edwardsii* (60% coverage) along with peptides that match part of the CRISP protein Catrin (gi 28972959) from *C. atrox*

Protein family	Taxonomy	gi	Sequence	charge	m/z	score
PI-Metalloproteinase	Agkistrodon contortrix laticinctus	1098019	QWVH <u>Q</u> IV <u>N</u> TINEIYR	2	958	34.93
PII-Metalloproteinase	Bothrops jararaca	123911605	LTTGSQCAEGLCCDQCK	1	1987.75	130
			DS <u>C</u> SCGANSCIMSATVSNEPSSR	1	2476.95	154
PII-Metalloproteinase	Echis ocelatus	320579329	YV <u>Q</u> LVIVADHSMVTK	1	1703.83	24.89
PII-Metalloproteinase	Crotalus adamanteus	338855314	YHFVANR	1	906.46	53
			NSVGIVQDHR	1	1124.59	48
PII-Metalloproteinase	Glodius saxatilis	31322301	YNSNLDTIR	2	548.28	54.37
PII-Metalloproteinase	Crotalos atrox	258618062	VALIGLEIWSSGELSK	1	1702.97	0.65
			YYTEVGEDCDCGPPANCQNPCCDAATCK	1	3312.36	80
PIII-metalloproteinase	Botrox atrox	205278807	SVGIVQDHGK	2	347.19	65.47
			HELGHNLGMDHDR	1	1546.71	94
			SYQFSDCSQNDHLR	1	1757.76	0.53
			YLISHTPQ <u>C</u> ILNEPLR	2	977.53	75.23
PIII-metalloproteinase	Trimeresurus gramineus	172044536	AQCGEGLCCDQCR	2	1613.61	0.32
			SCAGQSADCPTDDFHR	2	912.37	73.56
			AGEDCDCGSPANPCCDAATCK	1	2258.73	32.37
PIII-Metalloprotease	Glodius halys	4106007	AAGDTLEAFGDWR	2	704.84	63.91
			LRPGQQCAEGLCCDQCR	1	2107.9	0.54
PIII-Metalloprotease	Agkistrodon psicivorus	258618068	YLIDNRPPCILNK	2	808.45	41.89
	leucostoma		NLQGQGNFYCR	1	1356.66	46
PIII-Metalloproteinase	Atractaspis microlepidota andersoni	6007789	DTLDSFEEWR	2	649.3	52.74
PIII-metalloprotease	Bothrops jararaca	209870468	HDNAQLLTAIDF <u>N</u> GR	2	843.43	49.11
PIII-metalloprotease	Viridovipera stejnegeri	123900232	QLLTAIDFDGPTIGR	2	808.95	66.65
			LHSWVECESGECCEQCR	1	2226.88	0.61
PIII-Metalloproteinase	Trimeresurus flavoviridis	82217336	<u>Q</u> GNYYGY <u>C</u> R	1	1222.5	30.19
PIII-Metalloproteinase	Bothrops jararacussu	123889624	YSEDLDFGMVDHGTK	1	1729.7	36.49
PIII-Metalloproteinase	Echis coloratus	297593938	VTLNSFGEWR		1208.58	30.30
PIII-Metalloproteinase	Echis carinatus sochureki	297593788	LHSWVE <u>C</u> ESGE <u>CC</u> DQ <u>C</u> K		2183.82	38.67
PIII-Metalloproteinase	Bothrops jararaca	82219706	SECDIAESCTGQSADCPTDDFKR	1	2649.01	214
PIII-Metalloproteinase	Crotalus atrox	75570463	SECDIAESCTGQSADCPTDDFHR	1	2658.05	185
PIII-Metalloproteinase	Crotalus adamanteus	338855314	YEGDKTEICSR	1	1357.58	32
			MAHELGHNLGIDHDR	1	1714.76	51
PIII-Metalloproteinase	Trimeresurus flavoviridis	344925813	HSVGIVQDHGK	1	1176.59	44
PIII-Metalloproteinase	Crotalus adamanteus	338855316	LDVMVAVTMAHELAH	1	1636.80	125
PIII-Metalloproteinase	Crotalus adamanteus	338855326	YSEDLDYGMVDHGTK	1	1729.73	114
·			LFCKFNNFPCQYK	1	1766.79	72
			LHSWVECESGECCEQCK	1	2197.83	108
PIII-Metalloproteinase	Crotalus adamanteus	338855330	PKCILNEPLR	1	1239.65	64
·				1	2875.28	120
disintegrin	Gloydius shedaoensis	91680863	CTGQSAECPTDDFHR	2	890.86	57.51
			YFVEVGEECDCGLPAHC	1	2041.84	0.54
			SECDIAESCTGQSAECPTDDFHR	1	2672.07	185
disintegrin	Crotalus atrox	327507705	GDWNDDTCTGQSADCPR	1	1954.75	131
Serine proteinase	Crotalus adamanteus	338855332	AAYPEFGLPATSR	2	690.36	66.03
Serine proteinase	Bothrops jararaca	82233395	LDSPVSDSEHIAPL	2	740.38	42.25
	Domopo jararada	02200000		2		42.25

Table 1. Identification of proteins of the venom of S. c. edwardsii not predicted from the transcriptome analysis.

(Continued)



Table 1. (Continued)

Protein family	Taxonomy	gi	Sequence	charge	m/z	score
Serine proteinase	Viridovipera stejnegeri	82242793	IIGGDE <u>C</u> NIDEHR	2	764.36	56.68
C-type lectin	Sistrurus miliarius	21530567	GLQQGTNYHK	2	382.53	72.72
			F <u>C</u> SEQAEGGHLVSIESSEEAA	1	2239	0.54
			WSDGSSVSYENWIEAESK	1	2073.83	125
			D <u>C</u> PSGWSSYDQH <u>C</u> YR	1	1917.74	118
C-type lectin	Sistrurus miliarius	21530570	YDVWIGLR	2	511.28	60.77
			WSDGSSVNYENLIK	2	806.4	75.31
			DFD <u>C</u> PSDWYAYDQY <u>C</u> YR	1	2323.83	144
C-type lectin	Sistrurus miliarius	21530564	FTSMWIGLK	1	1082.57	64
			LASIHSSEEEAFVGK	1	1603.81	0.52
			TWDDAESF <u>C</u> YTQHR	1	1815.69	95
C-type lectin	Sistrurus miliarius	21530573	QNQYYVWIGLR	1	1439.75 1991.88	53
			ETEFLQWYNTD <u>C</u> EEK	1	1991.88	84
C-type lectin	Crotalus adamanteus	338855278	YEDWAEESYCVYFK	1	1888.79	92
C-type lectin	Crotalus durissus terrificus	82129809	WSDGSSVNYENLLK	2	806.40	75.31
			QNKYYVWIGLR	1	1439.75	46
			ETEFLQWYNTDCEEK	1	1991.86	75
L-amino acid oxidase	Bothrops neuwiedi pauloensis	195927838	GNPLEE <u>C</u> FR	2	561.26	59.52
			NGLSATSNPK	2	495.25	45.52
L-amino acid oxidase	Demansia vestigiata	118151720	YPVKPSEK	2	474.27	42.82
L-amino acid oxidase	Viridovipera stejnegeri	34014953	LSAAYVLAGAGHEVTVLEASER	1	2244.2	0.56
L-amino-acid oxidase	Naja kaouthia	124015192	QNDYEEFLEIAK	2	749.86	59.53
L-amino-acid oxidase	Crotalus atrox	124106294	TPYQFQHFSEALTAPFK	1	2012.03	80
CRISP	Crotalus atrox	28972959	EDKYTN <u>C</u> K	1	1057.43	53
			SLVQQAG <u>C</u> QDK	2	617.31	75.45
			MEWYPEAAANAER	1	1537.68	105
			SGPPCGDCPSACDNGLCTNPCTK	1	2525.08	152
CRISP	Vipera nikolskii	215262114	GNVDFDSESPR	1	1263.55	74
Venom nerve growth factor	Bothrops asper	186659795	NPNPVPTG <u>C</u> R	2	556.28	52.27
Venom nerve growth factor	Cryptophis nigrescens	123907150	HWNSY <u>C</u> TTT <u>Q</u> TFVK	1	1773.81	96
Phospholipase A2	Sistrurus catenatus tergeminus	45934756	LDTYTYSEENGEIICGGDDPCKK	1	2664.18	124

doi:10.1371/journal.pone.0092091.t001

[33]. Both proteins share about 87% sequence identity, with pronounced variations located primarily at the C-terminus. In addition, a peptide that matches a CRISP protein from another viperid species was also encountered.

The snake C-type lectin or C-type lectin-like protein families (snaclecs [34]) usually form disulfide linked homo- or hetero-dimers which are organized in oligomers to form larger quaternary protein complexes [35]. They affect the haemostatic system by interfering with coagulation factors or platelet activation [15]. The analysis of the proteome of *S. c. edwardsii* led to the identification of three isoforms of C-type lectins, as evidenced by the detection of sequence-specific unique peptides with sequence coverages between 67% (isoform 1) and 24% (isoform 3). Interestingly, however, the presence of peptide sequences identical to those of six C-type lectin proteins from the closely related rattlesnakes *Sistrurus miliarius*, *Crotalus adamanteus*, and *Crotalus terrificus* were also noted (Table 1). This brings the number of C-type

5			
4	5	MVLIRVLANLLILQLSYGQKSSELVIGGDECNINEHRSLALVYI	44
4 WULTRULAULUGISYACKSEUWGGECHINERESUWUP N 44 3 WULTRULAULUGISYACKSEUWGGECHINERESUWUP 44 4	9	MVLIRVLANLLILOLSYAOKSSELIIGGEECIINEHRFLVALYS	44
9 WLIRVLATLLIQUAXQKSSELVYGQBCWNERRESIVALTN 44 1 WLIRVLANLLIQUAXQKSSELVYGQBCWNERRESIVALTN 44 1	4		
3 WLIRVLANLLIGUSTACKSELVIGGOBCNWERHERVALM 44 1			
1 WILRVLANLLILGLSYAGKSELUVGGDECNIFEREFLAVVS 44 11 WILRVLANLLIGUSYAGKSELUVGGDECNIFEREFLAVVS 44 11			
2 WLIRVLANLLILQISYAOKSSELITGEDECNINERERLAVYS 44 6			
11 WULRVLANLILL_USYAQKSSELVVGCDECNINERRUVATIVY 44 10 WULRVLANLILLQLSYAQKSSELVIGGBECNINERRUVATIV4 44 17 MVSQVKANNILHRIEAMALIRLUVALILQLSYAQKSSELVIGGBECNINERRUVATIV4 44 18			
6			
10 WULTRULANLLILQLSVAQKSSELTIGGECNINERRULALVIS 44 7 MVSQVKANNLLHRIEKAMALIRLLVNLLILQLSVAQKSSELVIGGECNINERRULALVIS 60 5 TNGFLOCGTLINGEWULTAAHCDRKINLIFGUNGEDCONVERBERLALVIS 60 9 FKSKRFHCSGTLINGEWULTAAHCDRKINLIFGUNGEWURKAGUNGEDCONVERBERLALVIS 102 8 SSGFLCAGTLINGEWULTAAHCDEKNEOIGUGUNSKKVINEDDOGREPEKKEPICSSK 102 8 SSGFLCAGTLINGEWULTAAHCDEKNEOIGUNSKKVINEDDORPEKKEPICSSK 104 9 STGFCOGTLINGEWULTAAHCDEKNEOIGUNSKKVINEDDORPEKKEPICSSK 104 2 DGNQCOGTLINGEWULTAAHCDEKNEOIGUNSKKVINEDDORPEKKEPICSSK 104 10 SISKFFLCGGTLLNEEWULTAAHCNEKNEINTERLGUNSKVPNEKKORTVEPKEKEPICSSK 104 7 DKFQCGGTLINGEWULTAAHCNEWTENTIRGHINNNUTDDDORNEYFKKYFFRCHN 84 10 SISKFFLCGGTLLNEEWULTAAHCNEWTENTIRGHINNUTSTDDDURNEYKKYFFRCHN 84 11 DGIQGGGTLINGEWULTAAHCNEWTENTIRGHINNUTSTDDURTVEKKEFFCLSSK 104 12 NYTEWEKDINITALKERVINSKAUTVENUT ELGUNSKUTMAGUNSTENDUTTVAKEKYFFRCHN 84 13 SISKFFLCGGTLLNEEWULTAAHCNEWTENTIGGHINGEVUNTAKEKYFFRCHN 84 14 KDOVLDKDINLIKLDSVNSSTHIAPISLEGSPSTSVGSVCSUCHWUKGGTISSTKTIPOVEH 164 15 KNDEKKONINULIRLDSVNSSTHIAPISLSSPSSVGSVCSUCHWUKGGTISSTKTIPOV	11		
7 MYSQVKANALLERIEZMALIFILIVILLILUSYAQKSELVIGGDECVVNEHEFILALLYS 60 5 TNGFLOGGTLINQEWUTAAHCDRONGLIFFOVNELKGINEDQTENVEKEFICENEK 102 9 FKSKRECSGTLINQEWUTAAHCDRENNENIRIKIGUNSTKYNNEDQTENVEKEFICENEK 102 8 SSGFLCAGTLINQEWUTAAHCDSKNEWOLGUNSKKVINEDEQTENVEKEFICENEK 101 9 STGFFCAGTLINGEWUTAAHCDSKNEWOLGUNSKKVINEDEQTENVEKEFICLSSK 101 1 SSRTFLCGGTLINEEWUTAAHCDSKNEWOLGUNSKKVINEDEQTENVEKEFICLSSK 101 11 DEIQCGGTLINGEWUTAAHCDSKNEWICKGUNSKKVINEDEQTENVEKEFICLSSK 101 11 DEIQCGGTLINGEWUTAAHCDSKNEVISTURDDORRYFEKEFICLSSK 101 11 DEIQCGGTLINDEWUTAAHCDSKNEVISTURDDORRYFEKEFICLSSK 101 11 DEIQCGGTLINDEWUTAAHCDRYFUYILGUNSKVYNDDORRYFEKEFICLSSK 101 11 DEIQCGGTLINDEWUTAAHCDRYFURJELGUNSKVYNDDORRYFEKEFICLSSK 101 11 DEIQCGGTLINDEWUTAAHCDRYFURJELGUNSKVYNDDORRYFEKEFICLSSK 101 11 DEIQCGGTLINDEWUTAAHCDRYFURJELGUNSKVYNDDORRYFEKEFICLSSK 101 11 DKFOCGTLINDEWUTAAHCDRYFURJELGUNSKVYNDDORRYFEKEFICLSSK 101 11 DKFOCGTLINDEWUTAAHCDRYFERGUNSKVINTEDORRYFEKEKYFERCEN 102 12	6	SSELVIGGDECNINEHRFLVALYH	24
7 MYSQVKANALLERIEZMALIFILIVILLILUSYAQKSELVIGGDECVVNEHEFILALLYS 60 5 TNGFLOGGTLINQEWUTAAHCDRONGLIFFOVNELKGINEDQTENVEKEFICENEK 102 9 FKSKRECSGTLINQEWUTAAHCDRENNENIRIKIGUNSTKYNNEDQTENVEKEFICENEK 102 8 SSGFLCAGTLINQEWUTAAHCDSKNEWOLGUNSKKVINEDEQTENVEKEFICENEK 101 9 STGFFCAGTLINGEWUTAAHCDSKNEWOLGUNSKKVINEDEQTENVEKEFICLSSK 101 1 SSRTFLCGGTLINEEWUTAAHCDSKNEWOLGUNSKKVINEDEQTENVEKEFICLSSK 101 11 DEIQCGGTLINGEWUTAAHCDSKNEWICKGUNSKKVINEDEQTENVEKEFICLSSK 101 11 DEIQCGGTLINGEWUTAAHCDSKNEVISTURDDORRYFEKEFICLSSK 101 11 DEIQCGGTLINDEWUTAAHCDSKNEVISTURDDORRYFEKEFICLSSK 101 11 DEIQCGGTLINDEWUTAAHCDRYFUYILGUNSKVYNDDORRYFEKEFICLSSK 101 11 DEIQCGGTLINDEWUTAAHCDRYFURJELGUNSKVYNDDORRYFEKEFICLSSK 101 11 DEIQCGGTLINDEWUTAAHCDRYFURJELGUNSKVYNDDORRYFEKEFICLSSK 101 11 DEIQCGGTLINDEWUTAAHCDRYFURJELGUNSKVYNDDORRYFEKEFICLSSK 101 11 DKFOCGTLINDEWUTAAHCDRYFURJELGUNSKVYNDDORRYFEKEFICLSSK 101 11 DKFOCGTLINDEWUTAAHCDRYFERGUNSKVINTEDORRYFEKEKYFERCEN 102 12	10	MVLIRVLANLLILQLSYAQKSSEL IIGGEECNINEHR FLVALYH	44
5 TNGFLCGGTLINQEWULTAAHCDRGNNLIFFGUHRLKGLNKDAQTRVAKEKFICPNRK 102 9 FKSRKPHCSGTLINQEWULTAAHCDRKNITKLGTHSTKUTNEDUQTRVFKKFFICLSSK 101 8 SSGFLCAGTLINQEWULTAAHCDSKNFQNQLGVISKKVINEDEQTRDFKEKFICPNRK 102 3 STGFECAGTLINKEWULTAAHCDSKNFQNQLGVISKKVINEDEQTRDFKEKFICPNRK 101 3 STGFCCAGTLINKEWULTAAHCDSKNFQNQLGVISKKVNNEDEQTRDFKEKFICDSK 101 4 DENOCGGTLINEEWULTAAHCDSKNFQNGKLSKVPNNDGQTRVFKEKFICDSK 101 5 DENOCGGTLINEEWULTAAHCDSKNKVPNNDKGTRVFKEKFICDSK 101 6 SRSKTFLCGGTLINEEWULTAAHCDRKNKVFLTURUSKSTVPNKERFFCVSSK 101 7 DKCQGGTLINEEWULTAAHCDRKNVFNTRLGNISKVVPNDDDQRFYFKKKYFFCLSK 11 10 SRSKTFLCGGTLINEEWULTAAHCDRKNVFNTRLGNISKVVPNDDDDRTVXKEKYFFCRSK 14 7 DKFQCGGTLINEEWULTAAHCDRKNVFNTRLGNISKVVPNDDDRTVXKEKYFFCLSK 117 7 DKFQCGGTLINEEWULTAAHCDRKNVFNTRLGNISKVVNFDDDDRTVXKEKYFFCLSK 117 7 DKFQCGGTLINEEWULTAAHCDRKNVFNTRLGNISKVVNFDDDRTXXKEKYFFCLSK 117 7 DKFQCGGTLINEEWULTAAHCDRKNVFNTRLGNKNVFDDDDRTXXKEKYFFCLSK 117 7 DKFQCGGTLINEEWULTAAHCDRKNVFNTRLGNISKVYNTRUSKT 1162 8 TNTKWKKDDIMLIRLDNSYNSKIFLAFLSLSPSSPSVGSVCHIMGKGTSTKKKKYFVPNFN <th>7</th> <td></td> <td></td>	7		
9 FKSKREHCSGTLINÖEWUTAAHCNEKNIRIKUCTESTKUTNEDÜÖTKUVKKEFFCLSSK 104 4 STNFLCAGTLINÖEWUTAAHCDIKNFQIQLGVHSKKVINEDEGTRDPREKEFCCLSSK 102 3 STGFLCAGTLINÖEWUTAAHCDIKNFQIQLGVHSKKVSNEDEGTRDPREKEFCCLSSK 101 1 SRSKT-LCAGTLINÖEWUTAAHCDIKNFQIQLGVHSKKVSNEDEQTRDPREKEFCCLSSK 101 2 DENQCOGTLINEEWUTAAHCEOKNKLIGUNSKVVPNEDQUTKVFKERFCCLSSK 101 6 SRSKTFLCAGTLINEEWUTAAHCENKKKIHLGMHSKVVPNEDQUTKVFKERFCCLSSK 101 7 DENQCOGTLINEEWUTAAHCENKKHLIGUNSKVVPNEDQUTKVFKERFCCLSSK 101 6 SRSKTFLCAGTLINEEWUTAAHCNRVTVITLAGMHSKVVPNEDDDURTKEKKYFFCHS 84 7 DKFQCGGTLINEEWUTAAHCNRVTVITLAGMHNKNVFEDDDURTKAKKYFFCHS 84 7 DKFQCGGTLINEEWUTAAHCNRVTVITLGVMNISVKYDDEQRRFFKKKYFFCHS 84 7 DKFQCGGTLINEEWUTAAHCNRVTVITLGVMNISVKYDDEQRRFFKKKYFFCHS 84 7 DKFQCGGTLINEEWUTAAHCNRVTVITLGVMNISVKYDDEQRRFFKKKYFFCHS 84 8 TNNEWENDIMIRUNEEWUTAAHCNRVTVITLGVMNISVYDDEDQRFFKKKYFFCHS 84 9 TYTRNDKDIMLIRLDRVSSEHIAPLSLPSPSYGSVCHIMCWGATSFNTKKTFDVPH 10 NKDWKDIMLIRUDRVSSEHIAPLSLPSPSYGSVCRIMCWGUTSFFTKKTFDVPH 11 NYTKWDKDIMLIRUDRVSSEHIAPLSLPSSPSYGSVCRIMCWGUTSFFTKVILPDVH 12 NYTKWDKDIMLIRUDRVSSEHIAPLSLPSSPSYGSVCRIMCWGUTSFFTKVILPDVH 14 NYTKWDK		~ ~ ~	
9 FKSKREHCSGTLINÖEWUTAAHCNEKNIRIKUCTESTKUTNEDÜÖTKUVKKEFFCLSSK 104 4 STNFLCAGTLINÖEWUTAAHCDIKNFQIQLGVHSKKVINEDEGTRDPREKEFCCLSSK 102 3 STGFLCAGTLINÖEWUTAAHCDIKNFQIQLGVHSKKVSNEDEGTRDPREKEFCCLSSK 101 1 SRSKT-LCAGTLINÖEWUTAAHCDIKNFQIQLGVHSKKVSNEDEQTRDPREKEFCCLSSK 101 2 DENQCOGTLINEEWUTAAHCEOKNKLIGUNSKVVPNEDQUTKVFKERFCCLSSK 101 6 SRSKTFLCAGTLINEEWUTAAHCENKKKIHLGMHSKVVPNEDQUTKVFKERFCCLSSK 101 7 DENQCOGTLINEEWUTAAHCENKKHLIGUNSKVVPNEDQUTKVFKERFCCLSSK 101 6 SRSKTFLCAGTLINEEWUTAAHCNRVTVITLAGMHSKVVPNEDDDURTKEKKYFFCHS 84 7 DKFQCGGTLINEEWUTAAHCNRVTVITLAGMHNKNVFEDDDURTKAKKYFFCHS 84 7 DKFQCGGTLINEEWUTAAHCNRVTVITLGVMNISVKYDDEQRRFFKKKYFFCHS 84 7 DKFQCGGTLINEEWUTAAHCNRVTVITLGVMNISVKYDDEQRRFFKKKYFFCHS 84 7 DKFQCGGTLINEEWUTAAHCNRVTVITLGVMNISVKYDDEQRRFFKKKYFFCHS 84 8 TNNEWENDIMIRUNEEWUTAAHCNRVTVITLGVMNISVYDDEDQRFFKKKYFFCHS 84 9 TYTRNDKDIMLIRLDRVSSEHIAPLSLPSPSYGSVCHIMCWGATSFNTKKTFDVPH 10 NKDWKDIMLIRUDRVSSEHIAPLSLPSPSYGSVCRIMCWGUTSFFTKKTFDVPH 11 NYTKWDKDIMLIRUDRVSSEHIAPLSLPSSPSYGSVCRIMCWGUTSFFTKVILPDVH 12 NYTKWDKDIMLIRUDRVSSEHIAPLSLPSSPSYGSVCRIMCWGUTSFFTKVILPDVH 14 NYTKWDK	5	TNGFLCCCTLINOFWULTAAHCDRCNMLIFFCUHRLKCLNKDAOTRVAKEKFICPNRK	102
4 STNFLCAGTLIN0EWULTAAHCDSKNFOMQLGVHSKKVLNEDEQTRDREKKFICPNK 102 8 SSGFLCAGTLIN0EWULTAAHCDENNFOULGVHSKVNNEDEQTRDREKKFICPNK 101 1 STNFFCAGTLINNEWULTAAHCDENNFOULGVHSKVNNEDEQTRDREKKFICDSK 102 3 STNFCGGTLINNEWULTAAHCDENNKULTAHCDMENVENDEDQTRDREKKFICDSK 101 1 DSG0CGGTLINNEWULTAAHCDKNKLUGHUSKVNEDEQTRDREKKFICDSK 101 10 SSKTFLCGGTLINNEWULTAAHCDKNKLUGHUSKNKLOPTLBKKVFNEDDQTRVFKKKFICSSK 101 6 SSKTFLCGGTLINNEWULTAAHCDKNKLUTLGWINKNVENDEDQTRVFKKKFICSSK 101 7 DKFQCGGTLINNEWULTAAHCNKVFVIRLGWINKNVIEDDQTRVFKKKFICSSK 101 7 DKFQCGGTLINNEWULTAAHCNNEWIFLGWINNSVENDDQTRVFKKFFCSSK 101 7 DKFQCGGTLINNEWULTAAHCNNEHVEVILGWINNSVENDDQTRVFKKFFCSSK 101 7 DKFQCGGTLINNEWULTAAHCNNEHVEVILGWINNSVENDDQTRVFKKFFCSSK 104 7 DKFQCGGTLINNEWULTAAHCNNEHVEVILGWINNSVENDDQTRVFKKFFCSSK 104 7 DKFQCGGTLINNEWULTAAHCNNEHVEVILGWINNSVENDDQTRVFKKFFCSSK 104 8 TINNEWKINNEHI KLINPSVSSTHI APISLPSSPSVOSCHIMGWGAITSPHVLIPOVPH 162 9 TINRNKKINIELKLINPSVSSTHI APISLPSSPSVOSCHIMGWGAITSPHVLIPOVPH 162 10 NYTRNKKINIELKLINPSVSSTHI APISLPSSPSVOSCHIMGWGAITSPKKEIYPDVH 162 11 NYTRNKKINIELKLINPSVSSTHI APISLPSSPSVOSCHIMGWGAITSPKKEIYPDVH 164 11 NYTRNKKINIELKLNPSVSSTHI APISLPSSPSVOSCHIMGWGAITSPKKEIYPDVH			
8 SSGFLCAGTLINÚEWULTAAHCDINNÜÜÜLGVISKKVINBEDEÜTRPEKEFECLISK 101 3 STGFCCAGTLINEEWULTAAHCDSINFOMKUVISKUSMEDEÜTRPEKEFECLISK 101 2 DGNQCGGTLINEEWULTAAHCDKINKUSUFUSKUSMEDEÜTRPEKEFECLISK 104 2 DGNQCGGTLINEEWULTAAHCDKINKUSUFUSKUSKUSMEDEÜTRPEKEFECLISK 101 6 SRSKTFLOGGTLINEEWULTAAHCDKINKUSCHUSKUSKUSKUSKUSKUSKUSKUSKUSKUSKUSKUSKUSKU			
3 STGFFCAGTLINKEWULTAAHCDENNÖVÄLGVHEKKVONEDEÖTREKEKFFCED-K 101 1 SRSKTFLCGGTLINEEWULTAAHCDRINÖVÄLGVHEKKVONKUKOTREVEKKFFCUSSK 104 2 DONQCGGTLINDEWULTAAHCDEKNKKLOFGLISKVENKUKOTREVEKKFFCUSSK 101 11 DDIQCGGTLINDEWULTAAHCDEKNKKLOFGLISKVENKUKOTREVEKKFFCUSSK 101 6 SRSKTFLCGGTLINEEWULTAAHCDEKNKKLOFGLISKVENKUKOTREVEKKFFCUSSK 101 7 DKFQCGGTLINEEWULTAAHCDRVENVIRLGMINKUVENDDDDRREVEKKFFCUSSK 104 7 DKFQCGGTLINEEWULTAAHCDRVENVIRLGMINKUVIEDDDRREVEKKFFCUSSK 104 7 DKFQCGGTLINEEWULTAAHCDRVENVIRLGMINKUVIEDDDRREVEKKFFCUSSK 104 7 DKFQCGGTLINEEWULTAAHCDRVENVIRLGMINKUVIEDDDRREVEKKFFCUSSK 104 7 DKFQCGGTLINEEWULTAAHCDRVENVIRLGMINKUVIEDDDRREVEKKFFCUSSK 104 9 TUTRMKOKOIMLIKKEPUNSEH IAPLSLPSNPSUGSUCHMGWGAITSENVILPGVEH 162 9 TUTRMKOKOIMLIKKEPUNSEH IAPLSLPSSPSUGSUCHMGWGAITSENKTEPDOVEN 162 10 NTTRMKOKOIMLIKLENPUNSEH IAPLSLPSSPSUGSUCHMGWGAITSENKTENPUNENPUNE 11 NITLMOKOIMLIKLENPUNSEH IAPLSLPSSPSUGSUCHMGWGAITSENKTENPUNENPUNE 12 NTTRMKOKOIMLIKLENPUNSEH IAPLSLPSSPSUGSUCHMGWGAITSENKTENPUNENPUNE 14 NITLMOKOIMLIKLENPUNSEH IAPLSLPSSPSUGSUCHMGWGAITSENKTENPUNENPUNE 14 NITLMOKOIMLIKLENPUNENPUNENPUNENPUNENPUNENPUNENPUNENPU			
1 SRSKTFLCGGTLINEEWULTAAHCORRNWQIKLGMHSKYVENEDEQTRVEKKFFCLSSK 104 2 DGNQCGGTLINEEWULTAAHCORRNWQIKLGMHSKKYPENEDEQTRVEKKFFCLSSK 101 1 DGIQCGGTLINEEWULTAAHCORKNKILDGGLSKNVPNKDKQTRVPEKKFFCLSSK 101 6 SRSKTFLCGGTLINEEWULTAAHCORKNKILDGGLSKNVPNKDKQTRVPEKKFFCLSSK 101 7 DKFQCGGTLINEEWULTAAHCORKNULTULGMINKWUFDDDDHITYAKSKFFCLSSK 101 7 DKFQCGGTLINEEWULTAAHCORKNULTULGMINKWUFDDDHITYAKSKFFCLSSK 117 5 KNDEKKOLMLIRLKEWULTAAHCORKNULTULGMINKWUFDDDHITYAKSKFFCLSSK 117 6 SRSKTELCGGTLINEEWULTAAHCORKNULTULGMINKWUFDDDHITYAKSKFFCLSSK 117 7 DKFQCGGTLINEEWULTAAHCORKNULTULGMINKWUFDDDHITYAKSKFFFCLSSK 114 7 DKFQCGGTLINEEWULTAAHCORKNULTULGMINKWUFDDDHITYAKSKFFCLSSK 114 9 TTTRRDKOLMLIRLKEWUTAAHCONKUNTULGMINKWUFDDDHITYAKSKFFCLSSK 117 10 KDDEKKNIMLIRLKEWUNSEHIAPLSLPSSFSVOSUCHMGWGATSPEWUTLPOVEH 162 11 NYTKMKDKDIMLIRLNEDPVSSEHIAPLSLPSSFSVOSUCRIMGWGATSPETYEWPWH 161 11 NYTKMKDKDIMLIRLNEPVSYSEHIAPLSLPSSFSVOSUCRIMGWGATSPEKETYEWPWH 161 11 NYTKBOKDIMLIRLNEPVSYSEHIAPLSLPSSFSVOSUCRIMGWGATSPETENPVHLPVH 161 11 NYTKBOKDIMLIRLNEPVSYSEHIAPLSLPSSFSVOSUCRIMGWGATSPEKETPUPVH 161 11 NYTKBOKDIMLIRLNEPVSYSEHIAPLSLPSSFSVOSUCRIMGWGATSPEKETPUPVH 162 20 CANINILDPYCQAAYPSLTATSRTLCAGILEGGKSCDC			
2 DCNQCGGTLINDEW/LTAAHCDCKNK/LHLCMHSKKVPNKDKÖTRVPKEKEFCVSSK 101 11 DGIQCGGTLINDEW/LTAAHCDKKNKLOFGLHSKKVPNKDKÖTRVPKEKEFFCLSSK 101 6 SRSKTFLCGGTLINDEW/LTAAHCDKKNKLOFGLHSKKVPNKDKÖTRVPKEKEFFCLSSK 101 7 DKFQCGGTLINDEW/LTAAHCDKKNKLYIRLGMHKKNVIFDDDRAYPKKKYFFCKN 104 7 DKFQCGGTLINDEW/LTAAHCDKRHWIFLGWHKKNVIFDDDRAYPKKKYFFCKN 104 7 DKFQCGGTLINDEW/LTAAHCDKRHWIFLGWHKKNVIFDDDRAYPKKKYFFCLSSR 117 5 KNDEKDADIMLIRLKREVNISEHIAPLSLPSSPSVGSVCHIMGWGATTSPNVTLFGVPH 162 9 TYTKNDKDIMLIRLKREVNISEHIAPLSLPSSPSVGSVCRIMGWGTSPNVTLFGVPH 164 4 KDDKLKDIMLIRLNNPSVSSBHIAPLSLPSSPSVGSVCRIMGWGTSPTEKTIPPVPH 164 7 NYTWOKDIMLIRLDREVGNSEHIAPLSLPSSPSVGSVCRIMGWGTSPTEKTIPPVPH 164 1 NYTWOKDIMLIRLDREVGNSEHIAPLSLPSSPSVGSVCRIMGWGTSPCTEPVPH 164 11 NNKEWDKDIMLIRLNKPVSSENIAPLSLPSSPSVGSVCRIMGWGTSPCTEPVPH 164 12 NYTWOKDIMLIRLNKEVSSENIAPLSLPSSPSVGSVCRAMGWGTSPCTEPVPH 164 13 NNKEWDKDIMLIRLNKEVSSENIAPLSLPSSPSVGSVCRAMGWGTSPCTEPVPH 164 14 NNKEWDKDIMLIRLNKEVSSENIAPLSLPSSPSVGSVCRAMGWGTSPCTEPVPH 164 14 NNKEWDKDIMLIRLNKEVSSENIAPLSLPSSPSVGSVCRAMGWGTSPCTEPVPH 164 16 NFTKWDKDIMLIRLNKEVSSENIAPLSLPSSPSVGSVCRAMGWGTSPCTEPVPH 164 17 NYTKGDKDIMLIRLNKEVSSENIAPLSLSPSSPSVGSVCRAMGWGTSPCTEPVPH 164			
11 DCIQCGGTLINGEWWLTAAHCOKRYKLOFGLHSKNVPNENGNYPKEKFECLSSK 101 6 SRSKTFLCGGTLINEEWVLTAAHCONRVFLYIRLGMHINKNVTFDDDRINYAKEKFECHSSK 104 7 DKFQCGGTLINEEWVLTAAHCONRVFLYIRLGMHINKNVTFDDDRINYAKEKFFECKN 84 7 DKFQCGGTLINEEWVLTAAHCONRVFLYIRLGMHINKNVTFDDDRINYAKEKFFECKN 84 7 DKFQCGGTLINEEWVLTAAHCONRVHYIRLGMHINKNVTFDDDRINYAKEKFFECKN 84 7 DKFQCGGTLINEEWVLTAAHCONRVHYIRLGMHINKNVTFDDDRINYAKEKFFECKN 84 7 DKFQCGGTLINEEWVLTAAHCONRVHYIRLGMHINKNVTFDDDRINYAKEKFFECLSSK 104 9 TYTRWOKDIMLIRLDRYVNSEHIVPUNLPSNPSUGSVCHIMGGGITSPNVTLEGVPH 162 8 TNNEWEKDIMLIRLDRYVSSEHIAPLSLPSSPSVGSVCRIMGGGISSTSKETYPNVPH 162 1 NYTWEKDIMLIRLDRYVSSEHIAPLSLPSSPSVGSVCRIMGGGISSPNKETYPNVPH 164 1 NYTWEKDKDIMLIRLNERVYNSEHIAPLSLPSSPSVGSVCRIMGGGTSPCTLPDVPH 161 11 NNFKRENDIMLIRLNERVYSSEHIAPLSLPSSPSVGSVCRAMGGGTSPCTLPDVPH 164 10 NFTKRENDIMLIRLNERVYSSEHIAPLSLPSSPSVGSVCRAMGGGTSPCETLPDVPH 164 11 NNFKRENDIMLIRLNERVYSSEHIAPLSLPSSPSVGSVCRAMGGGTSPCETLPDVPH 164 12 NYTKØDKDIMLIRLNERVYSSEHIAPLSLPSSPSVGSVCRAMGGGTSPCETLPDVPH 164 13 NFTKEDKDIMLIRLNERVYSSEHIAPLSLPSSPSVGSVCRAMGGGTSPCETLPDVPH 164 14 NFTKEDKDIMLIRLNERVYSSEHIAPLSLPSSPSVGSVCRAMGGGOTSPCETLPDVH 164 15 CANINILDHEVCQAAYPSLTATSTRICAGILEGGKSSCDGDSGGPLI	1	SRSKTFLCGGTLLNEEWVLTAAHCDRRNMQIKLGMHSK TVPNEDEQTR VPKEKFFCLSSK	104
6 SRSTTPLCGGTLINEEWVLTAAHCNRVFLYTELCMHNENVNFDDDRYPEKKYFFFCRN 84 10 SRSKTPLCGGTLINEEWVLTAAHCNRVFNYTELGMHNENVTFDDDHIRYAKEKYFFFCNN 84 17 DKFQCGGTLINEEWVLTAAHCNRVFNYTELGMHNENVTDDQRRFFKKYFFCSNS 117 5 KNDEKOKDIMLIRLENEWVLTAAHCNRVFNYTELGVHNISVKYDDDQRRFFKKYFFCSNS 117 5 KNDEKOKDIMLIRLENEVVNLENNEHIVFVNLESNEPSLGSVCHIMGWGAITSPNVTLEGVH 162 9 TYTRWDKDIMLIRLENPVNSEHIAPLSLPSSPTSVGSVCRIMGWGINSPTKETYPDVH 162 8 TNNEWERDIMLIRLENPVNSEHIAPLSLPSSPTSVGSVCRIMGWGITSPTKEIYPNVFH 162 8 TNNEWERDIMLIRLENPVSNSEHIAPLSLPSSPPSVGSVCRIMGWGITSPTKEIYPNVFH 162 1 NYTFWDKDIMLIRLENPVSNSEHIAPLSLPSSPPSVGSVCRIMGWGITSPTKEIYPNVFH 164 11 NYTFWDKDIMLIRLENPVSNSEHIAPLSLPSSPPSVGSVCRIMGWGITSPTKEIYPNVFH 161 11 NNKWDKDIMLIRLENPVSSEHIAPLSLPSSPPSVGSVCRIMGWGITSPTKEIYPNVFH 161 11 NNKWDKDIMLIRLENKPVSSEHIAPLSLPSSPPSVGSVCRMMGWGITSPTKEIYPNVFH 164 10 NFTKRDKDIMLIRLENKPVSSEHIAPLSLPSSPPSVGSVCRMMGWGITSPTKEIYPNVFH 164 11 NNKGDKDIMLIRLENKPVSSEHIAPLSLPSSPPSVGSVCRMMGWGITSPTKEIYPNVFH 164 12 NYTKODKDIMLIRLENKPVSSEHIAPLSLPSSPSVGSVCRMMGWGITSPTKUTPDVFH 161 13 NKGDKDIMLIRLENKPVSSEHIAPLSLPSSPSVGSVCRMMGWGITSPTKEIYPNVFH 164 14 NOTKROKDIMLIRLENKPVSSENTALSLPSSPSVGSVCRMMGWGITSPTKUTPDVFH 161	2	DGNQCGGTLINEEWVLTAAHCEGNKMKIHLGMHSKKVPNKDKQTRVPKEKFFCVSSK	101
10 SRSKTFLCGGTLLNEEWVLTAAHCNRVFMYIRLGMHNKNVIFDDDÅIRYAKEKYFFRCRN 104 7 DKFQCGGTLINEEWVLTAAHCNRRHMYIRLGMHNKNVIFDDDÄIRYAKEKYFFRCRN 104 7 KNDEKOKDIMLIRLDSVNSSTHIAPISLPSNPSVGSVCHIMGWGAITSPNVTLEGVPH 162 9 TYTRØKDIMLIRLDRVSNSEHIAPISLPSNPSVGSVCHIMGWGAITSPNVTLEGVPH 164 4 KDDEKOKDIMLIRLDRVSNSEHIAPISLPSSPSVGSVCHIMGWGINSPTKETYPDVPH 162 8 TNNEWEKDIMLIRLDRVSNSEHIAPISLPSSPSVGSVCHIMGWGTISSTKETYPDVPH 162 9 NYTLWKKDIMLIRLDRVSNSEHIAPISLPSSPSVGSVCHIMGWGTISPTKVILPDVPH 161 10 NYTLWKKDIMLIRLDRVSNSEHIAPISLPSSPSVGSVCRIMGWGRISPTKTIPNVPH 164 2 NYTFWDKDIMLIRLDRVSNSEHIAPISLPSSPSVGSVCRIMGWGTISPTKVILPDVPH 161 11 NNKEWDKDIMLIRLNKVSYSEVIAPISLPSSPSVGSVCRVMGWGGTTSPDETIPDVPH 161 10 NTTKRKKDIMLIRLNKVSYSEVIAPISLPSSPSVGSVCRVMGWGGTTSPDETIPDVPH 164 7 NYTKGDKDIMLIRLNKVYSYSEVIAPISLPSSPSVGSVCRVMGWGTISPTKVILPDVPH 161 10 NTTKRDKOIMLIRLNKVYSYSEVIAPISLPSSPSVGSVCRVMGWGTISPTKVILPDVPH 164 7 NYTKGDKDIMLIRLNKVYSYSEVIAPISLPSSPSVGSVCRVMGWGTISPTKVILPDVPH 164 7 NYTKBDKDIMLIRLNKVYSYSEVIAPISLPSSPSVGSVCRVMGWGTISPTKVILPDVPH 164 7 NYTKBDKDIMLIRLNKVYSYSEVIAPISLPSSPSVGSVCRVMGWGTISPTKVILPDVPH 164 7 NYTKBDKDIMLIRLNKVYSYSEVIAPISLSPSPSVGSVCRVMGWGTISPTKVILPDVPH 164 7 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKSSCDDBSGPL	11	DGIQCGGTLINOEWMLTAAHCDGKRMKLOFGLHSKNVPNKDKOTRVPKEKFFCLSSK	101
10 SRSKTFLCGGTLLNEEWVLTAAHCNRVFMYIRLGMHNKNVIFDDDÅIRYAKEKYFFRCRN 104 7 DKFQCGGTLINEEWVLTAAHCNRRHMYIRLGMHNKNVIFDDDÄIRYAKEKYFFRCRN 104 7 KNDEKOKDIMLIRLDSVNSSTHIAPISLPSNPSVGSVCHIMGWGAITSPNVTLEGVPH 162 9 TYTRØKDIMLIRLDRVSNSEHIAPISLPSNPSVGSVCHIMGWGAITSPNVTLEGVPH 164 4 KDDEKOKDIMLIRLDRVSNSEHIAPISLPSSPSVGSVCHIMGWGINSPTKETYPDVPH 162 8 TNNEWEKDIMLIRLDRVSNSEHIAPISLPSSPSVGSVCHIMGWGTISSTKETYPDVPH 162 9 NYTLWKKDIMLIRLDRVSNSEHIAPISLPSSPSVGSVCHIMGWGTISPTKVILPDVPH 161 10 NYTLWKKDIMLIRLDRVSNSEHIAPISLPSSPSVGSVCRIMGWGRISPTKTIPNVPH 164 2 NYTFWDKDIMLIRLDRVSNSEHIAPISLPSSPSVGSVCRIMGWGTISPTKVILPDVPH 161 11 NNKEWDKDIMLIRLNKVSYSEVIAPISLPSSPSVGSVCRVMGWGGTTSPDETIPDVPH 161 10 NTTKRKKDIMLIRLNKVSYSEVIAPISLPSSPSVGSVCRVMGWGGTTSPDETIPDVPH 164 7 NYTKGDKDIMLIRLNKVYSYSEVIAPISLPSSPSVGSVCRVMGWGTISPTKVILPDVPH 161 10 NTTKRDKOIMLIRLNKVYSYSEVIAPISLPSSPSVGSVCRVMGWGTISPTKVILPDVPH 164 7 NYTKGDKDIMLIRLNKVYSYSEVIAPISLPSSPSVGSVCRVMGWGTISPTKVILPDVPH 164 7 NYTKBDKDIMLIRLNKVYSYSEVIAPISLPSSPSVGSVCRVMGWGTISPTKVILPDVPH 164 7 NYTKBDKDIMLIRLNKVYSYSEVIAPISLPSSPSVGSVCRVMGWGTISPTKVILPDVPH 164 7 NYTKBDKDIMLIRLNKVYSYSEVIAPISLSPSPSVGSVCRVMGWGTISPTKVILPDVPH 164 7 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKSSCDDBSGPL	6	SESKTFLCGGTLLNEEWVLTAAHCNEVFLYTELGMHNKNVNFDDDORRYPKKKYFFECHN	84
7 DKFQCGGTLINEEWVLTAAHCNRRHMYIFLGVHNISVKYDDEQRRFFKKKYFCLSSR 117 5 KNDEKOKDIMLIRLADSPVNSSTHIAPISLPSNPPSVGSVCHIMGWGAITSPNVTLPGVPH 162 9 TYTRRDKDIMLIRLARPVNNSEHIVPVNLPSNPPSUGSVCHIMGWGAITSPNVTLPGVPH 164 4 KDDEKOKDIMLIRLARPVNNSEHIVPVNLPSNPPSUGSVCRIMGWGAITSPNVTLPGVPH 162 8 TNNEWEKDIMLIRLANPVSNSHIAPLSLPSSPFSVGSVCRIMGWGAITSPNVTLPGVPH 162 9 TYTRUKDKDIMLIRLANPVSNSHIAPLSLPSSPPSVGSVCRIMGWGAITSPNKTTPDVPH 162 1 NYTEWDKDIMLIRLAPVSNSHIAPLSLPSSPPSVGSVCRIMGWGAITSPIKTTPDVPH 161 1 NYTEWDKDIMLIRLAPVSNSHIAPLSLPSSPPSVGSVCRIMGWGTISPTKETTPDVPH 161 11 NNKEWKDIMLIRLNKPVSVSEHIAPLSLPSSPPSVGSVCRMGWGTISPTKETTPDVPH 161 11 NNKEWKDIMLIRLNKPVSVSEHIAPLSLPSSPPSVGSVCRMGWGTISPDKTPDVPH 161 11 NNTKKDKDIMLIRLNKPVSVSEHIAPLSLPSSPPSVGSVCRMGWGTISPQETLPDVPH 164 10 NFTKMBKDIMLIRLNKPVSVSEHIAPLSLPSSPPSVGSVCRMGWGTISPQETLPDVPH 164 11 NNTKKDKDIMLIRLNKPVSVSEHIAPLSLPSSPPSVGSVCRMGWGTISPQETLPDVPH 164 12 NYTKGKDIMLIRLNKPVSVSEHIAPLSLPSSPPSVGSVCRMGWGTISPTKETYPDVPH 164 13 NYTKGKDIMLIRLNKPVSVSEHIAPLSLPSSPPSVGSVCRMGWGTISPQETLPDVPH 164 14 OANINILDRLAVCRAAPSLTATSRTLCAGILEGGKSCDGDSGPLICNGEJQGIVSW 220 5 CANINILDREVCQAAYPSLTATSRTLCAGILEGGKSCDGDSGPLICNGEJQGIVSW 222 2 CANINILDREVCRAAPSPMPMPTRTLCAGILEGGKDTCCMDSG			
5 KNDEKDKDIMLIRLEREPVINSETHIAPISLEPSNPPSVGSVCHIMGWGAITSPNVTLPGVPH 162 9 TYTRRDKDIMLIRLEREPVINSEHIAPISLEPSSPFSVGSVCHIMGWGAITSPNVTLPGVPH 164 4 KDDEKDKDIMLIRLENEPVSNSEHIAPISLEPSSPFSVGSVCHIMGWGINSFTKETYPDVPH 162 8 TNNEWEKDIMLIRLENEPVSNSEHIAPISLESSPPSVGSVCHIMGWGINSFTKETYPDVPH 161 1 NYTLMDKDIMLIRLENEPVSNSEHIAPISLESSPPSVGSVCHIMGWGINSFTKUTEPVPY 161 1 NYTLMDKDIMLIRLENEPVSNSEHIAPISLPSSPPSVGSVCRIMGWGRISPTKVILEPVPY 161 1 NNEEWEKDIMLIRLENEPVSNSEHIAPISLPSSPPSVGSVCRMGWGRISPTKVILEPVPY 161 1 NNEEWEKDIMLIRLENEPVSNSEHIAPISLPSSPPSVGSVCRMGWGRISPTKVILEPVPY 161 1 NNEEWEKDIMLIRLENEPVSNSEHIAPISLPSSPPSVGSVCRMGWGRISPTKVILEPVPY 161 10 NFTKRDKDIMLIRLENEPVSNSEHIAPISLPSSPPSVGSVCRVMGWGRISPTKVILEPVPH 161 11 NNEEWEKDIMLIRLENEPVSTSEVIAPISLEPSSPPSVGSVCRVMGWGRISPEKTUPDVPH 164 7 NYTKGDKDIMLIRLENEPVSTSEVIAPISLEPSSPPSVGSVCRVMGWGRISPEKTUPDVPH 164 7 NYTKGDKDIMLIRLENEPVSTSEVIAPISLEPSSPPSVGSVCRVMGWGRISPEKTUPDVPH 164 7 NYTKGDKDIMLIRLENEPVSTSEVIAPISLESSPPSVGSVCRVMGWGRISPEKTUPDVPH 164 7 NYTKGDKDIMLIRLENEPVSTSEVIAPISLESSPPSVGSVCRVMGWGRISPEKTUPDVPH 164 7 NYTKGDKDIMLIRLENEPVSTSEVIAPISLESSPPSVGSVCRVMGWGRISPEKTUPDVPH 164 7 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKSCODGSGPLICNGEIQGIVSW 220			
9 TYTRWDKDIMLIRLKRPVNNSEHIVPVNLPSNPESLGSVCHIMGWGATSPNVTLPGVPH 164 4 KDDEKDKDIMLIRLDRPVSNSEHIAPLSLPSSPTSVGSVCRIMGWGINSPTKETYPVPH 162 8 TNNEWEKDIMLIRLDRPVSNSAHIAPLSLPSSPFSVGSVCRIMGWGISPNKETYPVPH 162 1 NYTLWDKDIMLIRLDRPVSNSEHIAPLSLPSSPFSVGSVCRIMGWGRISPNKETYPVPH 164 1 NYTLWDKDIMLIRLDRPVSNSEHIAPLSLPSSPFSVGSVCRIMGWGISPNKETYPVPH 164 2 NYTWDKDIMLIRLDRPVSNSEHIAPLSLPSSPFSVGSVCRIMGWGISPNKETYPVPH 161 11 NNKEWKDIMLIRLNRPVSSEHIAPLSLPSSPFSVGSVCRIMGWGTSPCTEPVPPH 161 6 NFTKRDKDIMLIRLNRPVSSEHIAPLSLPSSPFSVGSVCRMGWGQTTSPQETLPDVPH 164 7 NYTKGDKDIMLIRLNRPVSSEHIAPLSLPSSPFSVGSVCRVMGWGGTTSPQETLPDVPH 164 7 NYTKGDKDIMLIRLNRPVSSSEHIAPLSLPSSPFSVGSVCRVMGWGGTTSPQETLPDVPH 164 10 NFTKRDKADMYSS SCANINILDLNEVCQAAYPSLTATSTLCAGILEGGKSCDGDSGGPLICNGETQGIVSW 220 24 CANINILDMEVCQAAYPSLTATSTLCAGILEGGKDSCKDOSGGPLICNGETQGIVSW 224 25 CANINILDMEVCQAAYPSLFATSTLCAGILEGGKDSCKDSGEGSGPLICNGETQGIVSW 224 26 CANINILDME	7	DRFQCGGILINEEWVLIAAHCNRRHMIIFLGVHNISVRIDDEQRRFFRRAIFCLSSR	
9 TYTRWDKDIMLIRLKRPVNNSEHIVPVNLPSNPESLGSVCHIMGWGATSPNVTLPGVPH 164 4 KDDEKDKDIMLIRLDRPVSNSEHIAPLSLPSSPTSVGSVCRIMGWGINSPTKETYPVPH 162 8 TNNEWEKDIMLIRLDRPVSNSAHIAPLSLPSSPFSVGSVCRIMGWGISPNKETYPVPH 162 1 NYTLWDKDIMLIRLDRPVSNSEHIAPLSLPSSPFSVGSVCRIMGWGRISPNKETYPVPH 164 1 NYTLWDKDIMLIRLDRPVSNSEHIAPLSLPSSPFSVGSVCRIMGWGISPNKETYPVPH 164 2 NYTWDKDIMLIRLDRPVSNSEHIAPLSLPSSPFSVGSVCRIMGWGISPNKETYPVPH 161 11 NNKEWKDIMLIRLNRPVSSEHIAPLSLPSSPFSVGSVCRIMGWGTSPCTEPVPPH 161 6 NFTKRDKDIMLIRLNRPVSSEHIAPLSLPSSPFSVGSVCRMGWGQTTSPQETLPDVPH 164 7 NYTKGDKDIMLIRLNRPVSSEHIAPLSLPSSPFSVGSVCRVMGWGGTTSPQETLPDVPH 164 7 NYTKGDKDIMLIRLNRPVSSSEHIAPLSLPSSPFSVGSVCRVMGWGGTTSPQETLPDVPH 164 10 NFTKRDKADMYSS SCANINILDLNEVCQAAYPSLTATSTLCAGILEGGKSCDGDSGGPLICNGETQGIVSW 220 24 CANINILDMEVCQAAYPSLTATSTLCAGILEGGKDSCKDOSGGPLICNGETQGIVSW 224 25 CANINILDMEVCQAAYPSLFATSTLCAGILEGGKDSCKDSGEGSGPLICNGETQGIVSW 224 26 CANINILDME	_		
4 KDDEKDKDIMLIRLDRPVSNSEHIAPLSLPSSPTSVGSVCRIMGWGINSPTKETYPDVPH 162 8 TNNEWERDIMLIRLDNPVSNSAHIAPLSLPSSPPSVGSUCRIMGWGTISSTKEIYPDVPH 162 3 KDDULDKDIMLIKLDSPVSNSEHIAPLSLPSSPPSVGSUCRIMGWGTISSTKEIYPDVPH 164 1 NYTLWDKDIMLIRLDRPVCNSEHIAPLSLPSSPPSVGSUCRIMGWGTISPTKEIYPDVPH 164 2 NYTFWDKDIMLIRLDRPVCNSEHIAPLSLPSSPSVGSUCRIMGWGTISPTKEIYPDVPH 164 11 NNKEWDKDIMLIRLNRPVNNSKHIAPLSLPSSPSDGSUCRIMGWGTSPTCPTVPVH 164 6 NFTKRDKDIMLIRLNRPVNSKHIAPLSLPSSPSDGSUCRIMGWGQTTSPQETLPDVPH 161 7 NYTKODKDIMLIRLNRPVSYSEHIAPLSLPSSPSVGSUCRVMGWGQTTSPQETLPDVPH 164 7 CANINLLDHEVCQAAYPSLTATSTLCAGILEGKSSCDGDSGPLICNCELQCIVSW 220 9 CANINLLDVECQAAYPSLTATSTLCAGILEGKSSCDGDSGPLICNCELQCIVSW 220 10 CANINLLDAVCCAAYPSLTATSTLCAGILEGKSCDCDSGGEDICNCELC			
8 TNNEWEKDIMLIRLNNPVSNSAHIAPLSLPSSPSVGSLCRIMGWGTISSTKEIYPNVPR 162 3 KDDVLDKDIMLIRLDSPVSNSEHIAPLSLPSSPSVGSVCRIMGWGTISSTKEIYPNVPR 161 1 NYTEWDKDIMLIRLDRPVGNSEHIAPLSLPSSPSVGSVCRIMGWGTISPTKVIPNVPR 164 2 NYTEWDKDIMLIRLNRPVSNSEHIAPLSLPSSPSVGSVCRIMGWGTISPTKVILPDVPH 161 11 NNKEWDKDIMLIRLNRPVSNSEHIAPLSLPSSPSVGSVCRIMGWGTISPTKEIYPDVPH 161 6 NFTKRDKDIMLIRLNRPVSYSEHIAPLSLPSSPSVGSVCRAMGWGTTSPQETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSYSEHIAPLSLPSSPSVGSVCRVMGWGTTSPQETLPDVPH 164 7 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKDSCGDSGGPLICNGEIQGIVSW 220 8 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKDSCGDSGGPLICNGEIQGIVSW 220 9 CANINILDAYCRAAYPEWPVTTRILCAGILEGGKDTCGDSGGPLICNGEIQGIVSW 221 10 CANINILLNYSVCRAAYPEYMLPATSRTLCAGILEGGKDTCGDSGGPLICNGEIQGILSW 222 11 CANINILLNYSVCRAAYPE-SIPATSRTLCAGILEGGKDTCGDSGGPLICNGEIQGILSW 222 7 CANINILDYEVCQAAYGSPATSRTLCAGILEGGKDTCGDSGGPLI	9	TYTRWDKDIMLIRLKRPVNNSEHIVPVNLPSNPPSLGSVCHIMGWGATTSPNVTLPGVPH	164
3 KDDVLDKDIMLIKLDSPVSNSEHIAPLSLPSSPSVGSVCHIIGWGSITPIEKTLPDVPY 161 1 NYTLBDKDIMLIKLDRPVSNSEHIAPLSLPSSPSVGSVCRIMGWGRISPKKTYPNVFH 164 2 NYTFWDKDIMLIKLDRPVSNSEHIAPLSLPSSPSVGSVCRIMGWGRISPTKEIYPDVPH 161 11 NNKEWDKDIMLIKLNRPVSNSEHIAPLSLPSSPSVGSVCRIMGWGTISPTKEIYPDVPH 161 6 NFTKRDKDIMLIKLNKPVSYSEHIAPLSLPSSPSVGSVCRAMGWGTISPTKEIYPDVPH 164 10 NFTKWDKDIMLIKLNKPVSYSEHIAPLSLPSSPSVGSVCRAMGWGTTSPQETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVRSYSEHIAPLSLPSSPSVGSVCRVMGWGTISPTGETUPDVPH 164 7 NYTKGDKDIMLIRLNKPVRSYSEHIAPLSLPSSPSVGSVCRVMGWGTISPSPCHUPVVP 177 5 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKSSCDGDSGGPLICNGEIQGIVSW 220 9 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKDSCGDSGGPLICNGEIQGIVSW 220 8 CANINILDAVCQRAYPSLTATSRTLCAGILEGGKDSCGDSGGPLICNGEIQGIVSW 220 8 CANINILDAVCQRAYPSLTATSRTLCAGILEGGKDSCGDSGGPLICNGEIQGIVSW 220 9 CANINILDAVCQRAYPELPATSRTLCAGILEGGKDTCGDSGGPLICNGEIQGIVSW 220 11 CANINILDAVCQRAYPELPATSRTLCAGILEGGKDSCGDSGGPLICNGEIQGIVSW 221 12 CANINILDAVCQRAYPELPATSRTLCAGILEGGKDTCGGSGGPLICNGQFQGIVSW 224 2 CVINILLNYSVCRAAYPEYQLPATSRTLCAGILEGGKDTCGGSGGPLICNGQFQGISW 224 2 CANINILDVEVCRAAYPEYGLPATSRTLCAGILEGGKDTCGGSGGPLICNGQFQGIVSW 224 10 CANINILDVEVCRAAYSSLPATSRTLCAGILEGGKDTCGDSG	4	KDDEKDKDIMLIRLDRPVSNSEHIAPLSLPSSPTSVGSVCRIMGWGINSPTKETYPDVPH	162
1 NYTLWDKDIMLIRLDRPVONSEHIAPLSLPSSPSVGSVCRIMGWGRISPNKETYPNVPH 164 2 NYTFWDKDIMLIRLDRPVONSEHIAPLSLPSSPPSVGSVCRIMGWGTISPTKVILDVPH 161 11 NNKEWDKDIMLIRLNRPVNNSKHIAPLSLPSSPPSVGSVCRIMGWGTISPTKVILDVPH 161 6 NFTKRDKDIMLIRLNRPVNSYSHIAPLSLPSSPPSVGSVCRMGWGTISPTKEIYPDVPH 164 10 NFTKRDKDIMLIRLNKPVSYSEVIAPLSLPSSPPSVGSVCRVMGWGTISPQETLPDVPH 144 10 NFTKRDKDIMLIRLNKPVSYSEVIAPLSLPSSPPSVGSVCRVMGWGTISPQETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSYSEVIAPLSLPSSPPSVGSVCRVMGWGTISPQETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSYSEVIAPLSLPSSPPSVGSVCRVMGWGTISPQETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSYSEVIAPLSLPSSPPSVGSVCRVMGWGTISPQETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSYSEVIAPLSLPSSPPSVGSVCRVMGWGTISPTKVMLPDVPH 177 5 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKSCDGDSGGPLICNGETQGIVSW 220 6 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKDTCGGDSGGPLICNGETQGIVSW 220 7 CANINILDAVCQAPYPSWPVTTRILCAGILEGGKDTCGGDSGGPLICNGETQGIVSW 220 8 CANINILDYAVCRAAYPELPATSRTLCAGILEGGKDTCCGDSGGPLICNGCPQGIVSW 224 11 CANINILDYEVCQAAYPE-LEASRVLCAGILEGGKDTCCGDSGGPLICNGCPQGINSW 224 2 CVNINILMYSVCRAAYPE-SLEASRVLCAGILEGGKDTCGGD	8	TNNEWEKDIMLIRLNNPVSNSAHIAPLSLPSSPPSVGSLCRIMGWGTISSTKEIYPNVPR	162
2 NYTFWDKDIMLIRLDRPVONSEHIAPLSLPSSPPSVGSVCRIMGWGTISPTKVILPDVPH 161 11 NNKEWDKDIMLIRLNRPVNSKHIAPLSLPSSPPSVGSVCRAMGWGQTTSPOTLPDVPH 164 10 NFTKRDKDIMLIRLNKPVSYSEHIAPLSLPSSPPSVGSVCRAMGWGQTTSPOETLPDVPH 144 10 NFTKWDKDIMLIRLNKPVSYSEHIAPLSLPSSPPSVGSVCRVMGWGQTTSPOETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSYSEHIAPLSLPSSPPSVGSVCRVMGWGQTTSPOETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSYSEHIAPLSLPSSPPSVGSVCRVMGWGTITSPNVMLPDVPY 177 5 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKSSCDGDSGGPLICNGEIQGIVSW 220 9 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKDSCCGDSGGPLICNGEIQGIVSW 220 8 CANINILDYAVCRAAYPWWPVTTRILCAGILEGGKDSCCGDSGGPLICNGEIQGIVSW 220 3 CANINILDYAVCRAAYPWWPVTTRILCAGILEGGKDSCCGDSGGPLICNGEIQGIVSW 220 3 CANINILDYEVCVAAYPEYLPATSRTLCAGILEGGKDSCCGDSGGPLICNGEIQGIVSW 220 3 CANINILDYEVCVAAYPEYME-MEATSRTLCAGILEGGKDSCCGDSGGPLICNGEFQGIVSW 221 11 CANINILDYEVCCAAYPEYEME-ATSRTLCAGILEGGKDTCCVGDSGGPLICNGOFQGIVSW 221 11 CANINILDYEVCCAAYSSLPATSRTLCAGILEGGKDTCCVGDSGGPLICNGOFQGIVSW 222 7 CANINILDYEVCAAYSSLPATSRTLCAGILEGGKDSCCGDSGGPLICNGOFQGIVSW 222 7 CANINILDYEVCAAYSSLPATSRTLCAGILEGGKDSCCGDSGGPLICNGOFQGIVSW 222 7 CANINILDYEVCAAYSSLPATSRTLCAGILEGGKDSCCGDSGGPLICNGOFQGIVSW 222 7 CANINILDYEVCAAYSSL	3	KDDVLDKDIMLIKLDSPVSNSEHIAPLSLPSSPPSVGSVCHIIGWGSITPIEKTLPDVPY	161
2 NYTFWDKDIMLIRLDRPVONSEHIAPLSLPSSPPSVGSVCRIMGWGTISPTKVILPDVPH 161 11 NNKEWDKDIMLIRLNRPVNSKHIAPLSLPSSPPSVGSVCRAMGWGQTTSPOTLPDVPH 164 10 NFTKRDKDIMLIRLNKPVSYSEHIAPLSLPSSPPSVGSVCRAMGWGQTTSPOETLPDVPH 144 10 NFTKWDKDIMLIRLNKPVSYSEHIAPLSLPSSPPSVGSVCRVMGWGQTTSPOETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSYSEHIAPLSLPSSPPSVGSVCRVMGWGQTTSPOETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSYSEHIAPLSLPSSPPSVGSVCRVMGWGTITSPNVMLPDVPY 177 5 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKSSCDGDSGGPLICNGEIQGIVSW 220 9 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKDSCCGDSGGPLICNGEIQGIVSW 220 8 CANINILDYAVCRAAYPWWPVTTRILCAGILEGGKDSCCGDSGGPLICNGEIQGIVSW 220 3 CANINILDYAVCRAAYPWWPVTTRILCAGILEGGKDSCCGDSGGPLICNGEIQGIVSW 220 3 CANINILDYEVCVAAYPEYLPATSRTLCAGILEGGKDSCCGDSGGPLICNGEIQGIVSW 220 3 CANINILDYEVCVAAYPEYME-MEATSRTLCAGILEGGKDSCCGDSGGPLICNGEFQGIVSW 221 11 CANINILDYEVCCAAYPEYEME-ATSRTLCAGILEGGKDTCCVGDSGGPLICNGOFQGIVSW 221 11 CANINILDYEVCCAAYSSLPATSRTLCAGILEGGKDTCCVGDSGGPLICNGOFQGIVSW 222 7 CANINILDYEVCAAYSSLPATSRTLCAGILEGGKDSCCGDSGGPLICNGOFQGIVSW 222 7 CANINILDYEVCAAYSSLPATSRTLCAGILEGGKDSCCGDSGGPLICNGOFQGIVSW 222 7 CANINILDYEVCAAYSSLPATSRTLCAGILEGGKDSCCGDSGGPLICNGOFQGIVSW 222 7 CANINILDYEVCAAYSSL	1	NYTLWDKDIMLIRLDRPVSNSEHIAPLSLPSSPPSVGSVCRIMGWGRISPNK ETYPNVPH	164
11 NNKEWDKDIMLIRLNRPVNNSKHIAPLSLPSKPPSQDTVCNIMGWGTISPTKEIYPDVPH 161 6 NFTKRDKDIMLIRLNKPVSYSEHIAPLSLPSSPPSVGSVCRAMGWGQTTSPQETLPDVPH 144 10 NFTKRDKDIMLIRLNKPVSYSEHIAPLSLPSSPPSVGSVCRVMGWGQTTSPQETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSYSEYIAPLSLPSSPPSVGSVCRVMGWGQTISPQETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSSEYIAPLSLPSSPPSVGSVCRVMGWGTISPQETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSSEYIAPLSLPSSPPSVGSVCRVMGWGTISPQETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSSEYIAPLSLPSSPSVGSVCRVMGWGTISPQETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSSEYIAPLSLPSSPSVGSVCRVMGWGTISPQETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSEYIAPLSLPSSPSVGSVCRVMGWGTISPQETLPDVPH 164 7 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKSCDGDSGGPLICNGETGGIVSW 220 8 CANINILDAVCQAAYPSLTATSRTLCAGILEGGKDSCQEDSGGPLICNGEIQGIVSW 220 9 CANINILDDAVCCPPYEPLPATSRTLCAGILEGGKDSCKDSGGPLICNGEIQGIVSW 220 11 CANINILDAVCCRAAYPSWPVTTRILCAGILEGGKDTCVDSGGPLICNGEPGGIVSW 221 2 CVININLINYSVCRAAYPEYDEATSRTLCAGILEGGKDTCVDSGSGPLICNGEPGGIVSW 221 2 CVININLLDAVCCRAAYSSLPATSRILCAGILEGGKDTCVDSGSGPLICNGEPGGIVSW 221 11 CANINILDAVCCRAAYSSLPATSRILCAGILEGGKDTCMDSGGP			
6 NFTKRDKDIMLIRLNKPVSYSEHIAPLSLPSSPPSVGSVCRAMGWGQTTSPQETLPDVPH 144 10 NFTKWDKDIMLIRLNKPVSYSEYIAPLSLPSSPPSVGSVCRVMGWGQTTSPQETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSYSEYIAPLSLPSSPPSVGSVCRVMGWGTTSPQETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVSYSEYIAPLSLPSSPPSVGSVCRVMGWGTTSPQETLPDVPH 164 7 NYTKGDKDIMLIRLNKPVRKSEHIAPLSLPSSPPSVGSVCRVMGWGTTSPQETLPDVPH 164 7 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKSSCDGDSGGPLICNGEIQGIVSW 220 9 CANINILHEVCQAAYPSLTATSRTLCAGILEGGKDACGGDSGGPLICNGEIQGIVSW 220 8 CANINILDDAVCQPAYPSLPATSRTLCAGILEGGKDSCQGDSGGPLICNGEIQGIVSW 220 3 CANINILDDAVCQPAYPSLPATSRTLCAGILEGGKDCGGDSGGPLICNGEIQGIVSW 220 3 CANINILDDAVCQAAYPEYMLPATSRTLCAGILEGGKDTCGDSGGPLICNGQFQGIVSW 220 3 CANINILDVEVCAAAYPEYMLPATSRTLCAGILEGGKDTCGGDSGGPLICNGQFQGIVSW 221 11 CANINILDYEVCAAAYPEYMEYS-SLPATSRTLCAGILEGGKDTCGGDSGGPLICNGQFQGIVSW 224 2 CVNINILLNYSVCRAAYPEY-SLPATSRTLCAGILEGGKDTCMPDSGGPLICNGQFQGIVSW 224 10 CANINILDYEVCRAAYGSSPATSRTLCAGILEGGKDTCMPDSGGPLICNGQFQGIVSW 222 7 CANINILDYEVCRAAYGGSPATSRTLCAGILEGGKDTCMPDSGGPLICNGQFQGIVSW 222 7 CANINILDYEVCRAAYGGSPATSRTLCAGILEGGKDTCMPDSGGPLICNGQFQGIVSW 222 8 GAHPCGQLKPGGVTKVFDYDVDWIQSIIAGNTDATCPP 258 9 GPDTCAQPDKPSLYKVFDYDUNUQSIIAGNTDATCPP 257 <th></th> <td></td> <td></td>			
10NFTKWDKDIMLIRLNKPVSYSEYIAPLSLPSSPSVGSVCRVMGWGQTTSPQETLPDVPH1647NYTKGDKDIMLIRLNKPVRKSEHIAPLSLPSSPSVGSVCRVMGWGTITSPNVMLPDVPH1775CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKSSCDGDSGGPLICNGEIQGIVSW2209CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKSSCDGDSGGPLICNGEIQGIVSW2224CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKDACGGDSGGPLICNGEIQGIVSW2203CANINILDYAVCRAAYPWMPVTTRILCAGILEGGKDACGGDSGGPLICNGEIQGIVSW2203CANINILDDAVCQPYFELPATSRTLCAGILEGGKDTCGGDSGGPLICNGEIQGIVSW2204CANINILDDAVCQAAYPWMPVTTRILCAGILEGGKDTCGGDSGGPLICNGEIQGIVSW2205CANINILDYAVCRAAYPWMPVTTRILCAGILEGGKDTCGGDSGGPLICNGGPQGIVSW2206CANINILDYEVCVAAYPEYHLPATSRTLCAGILEGGKDTCGGDSGGPLICNGGPQGIVSW22111CANINILDQAVCREIYPGLLEKSRVLCAGILEGGKDTCVGDSGGPLICNGGPQGISW22111CANINILDYEVCRAAHPSGIPATSRILCAGILEGGKDTCGGDSGGPLICNGGPQGIVSW2227CANINILDYEVCRAAHPSGIPATSRILCAGILEGGKDTCGGDSGGPLICNGQFQGIVSW2235GGDICAQPREPGVYTKVFDYTDWIQNIIAGNTDATCPP2589GPDTCAQPDKPSLYTKVPDHJUNIQSIIAGNTDATCPP2588GAHPCGRALNPGFYTKVPDYIDWIQSIIAGNTDATCPP2571GDDPCQVLQPALYTGVFDHLDWIQSIIAGNTDATCPP2576GDDCAQSHVPALYSKVFDHLDWIQSIIAGNTDATCPP2576GDGCAQPDKPALYSKVFDHLDWIQSIIAGNTDATCPP2576GDGCAQPDKPALYSKVFDHLDWIQSIIAGNTDATCPP2576GDGCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQSIISGNTDATCPP24310GDGSCAQPDKPALYSKVFDHLDWI			
7 NYTKGDKDIMLIRLNKPVRKSEHIAPLSLPSSPSVGSVCRVMGWGTITSPNVMLPDVPY 177 5 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKSSCDGDSGGPLICNGEIQGIVSW 220 9 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKDSCQDSGGPLICNGEFQGIVFW 222 4 CANINILDHEVCQAAYPSLTATSGTLCAGILEGGKDSCQGDSGGPLICNGEFQGIVFW 222 4 CANINILDHEVCQAAYPSLTATSGTLCAGILEGGKDSCQGDSGGPLICNGEFQGIVSW 220 8 CANINILDAVCQPPYFELEATSRTLCAGILEGGKDSCQGDSGGPLICNGEFQGIVSW 220 3 CANINILDAVCQPPYFELEATSRTLCAGILEGGKDCGGDSGGPLICNGQFQGIVSW 220 3 CANINILDYSVCRAAYPEYGLPATSRTLCAGILEGGKDTCVGDSGGPLICNGQFQGIVSW 221 11 CANINILDQAVCREIYPGLEATSRTLCAGILEGGKDTCVGDSGGPLICNGQFQGINSW 221 11 CANINILDYSVCRAAYPEYGLPATSRTLCAGILEGGKDTCWGDSGGPLICNGQFQGINSW 221 11 CANINILDQAVCREIYPGLEATSRTLCAGILEGGKDTCMRDSGGPLICNGQFQGINSW 221 11 CANINILDYEVCRAAHPOFGLPATSRTLCAGILEGGKDTCMRDSGGPLICNGQFQGINSW 222 7 CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW 222 7 CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW 235 5 GGDICAQPREPGVYTKVFDYTDWIQNIIAGNTDATCPP 258 9 GPDTCAQPDKPSLYTKVFDYIDWIQSIIAGNTAATCPP 257 1 GDDPCGY VLQPALYTGVFDHLDWIQSIIAGNTDATCPP 257 1 GDDPCAQPKPALYSKVFDHLDWIQNIIAGNTTNACPP 257 6 <td< th=""><th></th><td></td><td></td></td<>			
5 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKSSCDGDSGGPLICNGEIQGIVSW 220 9 CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKSSCDGDSGGPLICNGEFQGIVFW 222 4 CANINILHHEVCQAAYPSLTATSGTLCAGILEGGKDACGGDSGGPLICNGEIQGIVSG 220 8 CANINILDYAVCRAAYPSMPVTTRILCAGILEGGKDSCQGDSGGPLICNGEIQGIVSW 220 3 CANINILDYAVCRAAYPSMPVTTRILCAGILEGGKDSCQGDSGGPLICNGEIQGIVSW 220 3 CANINILDYAVCRAAYPEYMPVTTRILCAGILEGGKDSCGGDSGGPLICNGEIQGIVSW 220 3 CANINILDYAVCRAAYPEYMLPATSRTLCAGILEGGKDSCGEDSGGPLICNGEIQGIVSW 224 2 CVNINLLNYSVCRAAYPEYGLPATSRTLCAGILEGGKDTCVGDSGGPLICNGQFQGILSW 224 2 CVNINLLDYSVCRAAYPEYGLPATSRTLCAGILEGGKDTCVGDSGGPLICNGQFQGILSW 224 11 CANINLLDYEVCRAAYS-SLPATSRVLCAGILEGGKDTCVGDSGGPLICNGQFQGILSW 221 11 CANINLLDYEVCRAAYSGSPATSRTLCAGILEGGKDTCMPDSGPLICNGQFQGILSW 222 7 CANINLLDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW 222 7 CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW 235 5 GGDICAQPREPGVYTKVFDYDWIQNIIAGNTDATCPP 258 8 GAHPCGQLKPGVYTKVFDYIDWIQSIIAGNTDATCPP 258 8 GAHPCGQALKPGVYTKVFDYNDWIQSIIAGNTDATCPP 257 11 GDDPCGYVLQPALYGVFDHLDWIQSIIAGNTDATCPP 257 12 GDPCAQSHVPALYIKVFDYTEWIQSIIAGNTDATCPP 257 13 GDPCAQPHPALYSKVFDHLD			
9CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKSSCDGDSGGPLICNGEFQGIVFW 2224CANINILHHEVCQAAYPSLTATSGTLCAGILEGGKDCGGDSGGPLICNGEIQGIVSG 2208CANINILDYAVCRAAYPWWPVTTRILCAGILEGGKDCQGDSGGPLICNGEIQGIVSG 2203CANINILDDAVCQPPYPELPATSRTLCAGILEGGKDCQGDSGGPLICNGEIQGIVSG 2203CANINILDDAVCQPPYPELPATSRTLCAGILEGGKDCQGDSGGPLICNGEIQGIVSG 2204CANINILDDAVCQPPYPELPATSRTLCAGILEGGKDCQGDSGGPLICNGCPQGIVSW 2205CANINILNYSVCRAAYPEYMLPATSRTLCAGILEGGKDSCKGDSGGPLICNGCPQGILSW 2242CVNINLLNYSVCRAAYPEYGLPATSRTLCAGILEGGKDTCVGDSGGPLICNGCPQGILSW 22411CANINILDQAVCREIYPGLLEKSRVLCAGILEGGKDTCVGDSGGPLICNGCPQGINSW 22410CANINILDYEVCRAAYSSLPATSRVLCAGULEGGIDTCHRDSGGPLICNGCPQGIVSW 20410CANINILDYEVCRAAYSSLPATSRVLCAGULEGGIDTCHRDSGGPLICNGCPQGIVSW 2227CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGCPQGIVSW 2235GGDICAQPREPGVYTKVFDYTDWIQNIIAGNTDATCPP 2589GPDTCAQPDKPSLYKVFDYLDWIQSIIAGNTDATCPP 2589GPDTCAQPDKPSLYKVFDYDDWIQSIIAGNTDATCPP 2588GAHPCGQULTPGFYTTVFDVIDWIQSIIAGNTDATCPP 2588GAHPCGQALKPGYYKVFDYNDWIQSIIAGNTDATCPP 2571GDDPCAYUQPALYGVFDHLDWIQSIIAGNTDATCPP 2572GSPNCGYVREPGLYKVFDHLDWIQSIIAGNTDATCPP 2576GDGSCAQPDKPALYSKVFDHLDWIQSIIAGNTTVNCPP 24210GDDSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP 24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP 273	/	NYTKGDKDIMLIRLNKPVRKSEHIAPLSLPSSPPSVGSVCRVMGWGTITSPNVMLPDVPY	T / /
9CANINILDHEVCQAAYPSLTATSRTLCAGILEGGKSSCDGDSGGPLICNGEFQGIVFW 2224CANINILHHEVCQAAYPSLTATSGTLCAGILEGGKDCGGDSGGPLICNGEIQGIVSG 2208CANINILDYAVCRAAYPWWPVTTRILCAGILEGGKDCQGDSGGPLICNGEIQGIVSG 2203CANINILDDAVCQPPYPELPATSRTLCAGILEGGKDCQGDSGGPLICNGEIQGIVSG 2203CANINILDDAVCQPPYPELPATSRTLCAGILEGGKDCQGDSGGPLICNGEIQGIVSG 2204CANINILDDAVCQPPYPELPATSRTLCAGILEGGKDCQGDSGGPLICNGCPQGIVSW 2205CANINILNYSVCRAAYPEYMLPATSRTLCAGILEGGKDSCKGDSGGPLICNGCPQGILSW 2242CVNINLLNYSVCRAAYPEYGLPATSRTLCAGILEGGKDTCVGDSGGPLICNGCPQGILSW 22411CANINILDQAVCREIYPGLLEKSRVLCAGILEGGKDTCVGDSGGPLICNGCPQGINSW 22410CANINILDYEVCRAAYSSLPATSRVLCAGULEGGIDTCHRDSGGPLICNGCPQGIVSW 20410CANINILDYEVCRAAYSSLPATSRVLCAGULEGGIDTCHRDSGGPLICNGCPQGIVSW 2227CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGCPQGIVSW 2235GGDICAQPREPGVYTKVFDYTDWIQNIIAGNTDATCPP 2589GPDTCAQPDKPSLYKVFDYLDWIQSIIAGNTDATCPP 2589GPDTCAQPDKPSLYKVFDYDDWIQSIIAGNTDATCPP 2588GAHPCGQULTPGFYTTVFDVIDWIQSIIAGNTDATCPP 2588GAHPCGQALKPGYYKVFDYNDWIQSIIAGNTDATCPP 2571GDDPCAYUQPALYGVFDHLDWIQSIIAGNTDATCPP 2572GSPNCGYVREPGLYKVFDHLDWIQSIIAGNTDATCPP 2576GDGSCAQPDKPALYSKVFDHLDWIQSIIAGNTTVNCPP 24210GDDSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP 24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP 273			
4CANINILHHEVCQAAYPSLTATSGTLCAGILEGGKDACGGDSGGPLICNGEIQGIVSG2208CANINILDYAVCRAAYPWWPVTTRILCAGILEGGKDSCQGDSGGPLICNGEIQGIVSW2203CANINILDZAVCRAAYPWWPVTTRILCAGILEGGKDSCQGDSGGPLICNGOFQGIVSW2101CANINILDDAVCQPPYPELPATSRTLCAGILEGGKDTCGGDSGGPLICNGQFQGIASW2242CVNINLLNYSVCRAAYPEYGLPATSRTLCAGILEGGKDTCGGDSGGPLICNGQFQGIASW22111CANINILDQAVCREIYPGLEKSRVLCAGILEGGKDTCGGDSGGPLICNGQFQGIASW2296CANINILDYEVCRAAYPE-SLPATSRILCAGVLEGGIDTCHRDSGGPLICNGQFQGIVSW20410CANINLLDYEVCRAAYSSLPATSRVLCAGILEGGKDSCCGDSGGPLICNGQFQGIVSW2227CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW2355GGDICAQPREPGVYTKVFDYTDWIQNIIAGNTDATCPP2589GPDTCAQPDKPSLYTKVFDHLDWIQSIIAGNKTVNCPP2604GEHPCGQVLTPGFYTTVFDYIDWIQSIIAGNTDATCPP2583GAHPCGRRLNPGFYTKVFDYIDWIQSIIAGNTDATCPP2571GDDPCGYVLQPALYIGVFDHLDWIQSIIAGNTDATCPP2572GSPNCGYVREPGLYTKVFDHLDWIQSIIAGNTDATCPP25911GGDPCAQSHVPALYIKVFDHLDWIQSIIAGNTDATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQSIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP270	5	CANINILDHEVCQAAYPSLTATSRTLCAGILEGGK SSCDGDSGGPLICNGEIQGIVSW	220
4CANINILHHEVCQAAYPSLTATSGTLCAGILEGGKDACGGDSGGPLICNGEIQGIVSG2208CANINILDYAVCRAAYPWWPVTTRILCAGILEGGKDSCQGDSGGPLICNGEIQGIVSW2203CANINILDZAVCRAAYPWWPVTTRILCAGILEGGKDSCQGDSGGPLICNGQFQGIVSW2203CANINILDDAVCQPPYPELPATSRILCAGILEGGKDTCGGDSGGPLICNGQFQGIVSW2191CANINILNYSVCRAAYPEYGLPATSRILCAGILEGGKDTCVGDSGGPLICNGQFQGIVSW2212CVNINLLNYSVCRAAYPEYGLPATSRILCAGILEGGKDTCVGDSGGPLICNGQFQGIVSW22111CANINILDQAVCREIYPGLEKSRVLCAGILEGGKDTCGGDSGGPLICNGQFQGIVSW2196CANINILDYEVCRAAYSSLPATSRULCAGULEGGIDTCHRDSGGPLICNGQFQGIVSW20410CANINILDYEVCRAAYSSLPATSRULCAGULEGGIDTCHRDSGGPLICNGQFQGIVSW2227CANINILDYEVCQAAYGGSPATSRILCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW2355GGDICAQPREPGVYTKVFDYTDWIQNIIAGNTDATCPP2589GPDTCAQPDKPSLYTKVFDHLDWIQSIIAGNKTVNCPP2604GEHPCGQULTPGFYTTVFDYIDWIQSIIAGNTDATCPP2571GDDPCGYVLQPALYIGVFDHLDWIQSIIAGNTDATCPP25911GGDPCAQSHVPALYIKVFDHLDWIQSIIAGNTDATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQSIIAGNTDATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP273	9	CANINILDHEVCOAAYPSLTATSRTLCAGILEGGKSSCDGDSGGPLICNGEFOGIVFW	222
8 CANINILDYAVCRAAYPWWPVTTRILCAGILEGGKDSCOGDSGGPLICNGEIQGIVSW 220 3 CANINILDDAVCQPPYPELPATSRTLCAGILEGGKDCGGDSGGPLICNGQFQGIVFY 219 1 CANINILNYEVCVAAYPEYMLPATSRTLCAGILEGGKDCGDSGGPLICNGQFQGILSW 224 2 CVNINLLNYSVCRAAYPEYMLPATSRTLCAGILEGGKDTCVGDSGGPLICNGQFQGIASW 221 11 CANINILDQAVCREIYPGLLEKSRVLCAGILEGGKDTCVGDSGGPLICNGQFQGIVSW 221 6 CANINLLDYEVCRAAYE-SLPATSRTLCAGILEGGKDTCGGDSGGPLICNGQFQGIVSW 221 10 CANINLLDYEVCRAAYE-SLPATSRVLCAGVLEGGIDTCHRDSGGPLICNGQFQGIVSW 204 10 CANINILDYEVCRAAYE-SLPATSRVLCAGVLEGGIDTCNRDSGGPLICDQFQGIVSW 222 7 CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW 222 7 CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW 222 7 CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW 223 5 GGDICAQPREPGVYTKVFDYTDWIQNIIAGNTDATCPP 258 9 GPDTCAQPDKPSLYTKVFDYIDWIQSIIAGNTDATCPP 258 8 GAHPCGRALNPGFYTKVFDYIDWIQSIIAGNTATCPP 257 1 GDDPCGYULQPALYIGVFDHLDWIQSIIAGNTAATCPP 257 1 GDPCAQSHVPALYSKVFDHLDWIQSIIAGNTAATCPP 257 6 GDSCAQPDKPALYSKVFDHLDWIQSIIAGNTATCPP 257 6 GDSCAQPDKPALYSKVFDHLDWIQSIIAGNTATCPP 257 6 GDSCAQPDKPALYSKVFDHLDWIQSIIAGNTATCPP 257 <td< th=""><th></th><td>CANINILHHEVCOAAYPSLTATSGTLCAGILEGGKDACGGDSGGPLJCNGETOGIVSG</td><td>220</td></td<>		CANINILHHEVCOAAYPSLTATSGTLCAGILEGGKDACGGDSGGPLJCNGETOGIVSG	220
3CANTNLLDDAVCQPPYPELPATSRTLCAGILEGGKDTCGGDSGGPLICNGQFQGIVFY2191CANINILNYEVCVAAYPEYMLPATSRTLCAGILEGGKDTCGGDSGGPLICNGQFQGILSW2242CVNINLLNYSVCRAAYPEYGLPATSRTLCAGILEGGKDTCVGDSGGPLICNGQFQGILSW22111CANINILDQAVCREIYPGLLEKSRUCCAGILEGGKDTCGGDSGGPLICNGEIQGILSV2196CANINLLDYEVCRAAHPQFGLPATSRILCAGVLEGGIDTCHRDSGGPLICNGQFQGIVSW20410CANINLLDYEVCRAAHSSLPATSRVLCAGVLEGGIDTCHRDSGGPLICNGQFQGIVSW2227CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW2235GGDICAQPREPGVYTKVFDYTDWIQNIIAGNTDATCPP2589GFDTCAQPDKPSLITKVFDHLDWIQSIIAGNTDATCPP2588GAHPCGRRLNPGFYTKVFDYIDWIQSIIAGNTDATCPP2583GAHPCGQALKPGVYTKVFDYNDWIQSIIAGNTDATCPP2571GDDPCGYVLQPALYGVFDHLDWIQSIIAGNTDATCPP2571GDDPCGYVLQPALYGVFDHLDWIQSIIAGNTDATCPP25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQSIIAGNTDATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP273			
1CANINILNYEVCVAAYPEYMLPATSRTLCAGILEGGKDSCKGDSGGPLICNGQFQGILSW 2242CVNINLLNYSVCRAAYPEYGLPATSRTLCAGILEGGKDTCVGDSGGPLICNGQFQGIASW 22111CANINILDQAVCREIYPGLLEKSRVLCAGILEGGKDTCGGDSGGPLICNGFQGIASW 2216CANINLLDYEVCRAAHPQFGLPATSRILCAGVLEGGIDTCHRDSGGPLICNGQFQGIVSW 20410CANINLLDYEVCRAAYSSLPATSRVLCAGVLEGGIDTCHRDSGGPLICNGQFQGIVSW 2227CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW 22355GGDICAQPREPGVYTKVFDYTDWIQNIIAGNTDATCPP 2589GPDTCAQPDKPSLYTKVFDHLDWIQSIIAGNTDATCPP 2604GEHPCGQVLTPGFYTTVFDYIDWIQSIIAGNTDATCPP 2588GAHPCGRRLNPGFYTTVFDYIDWIQSIIAGNTATCPP 2583GAHPCGQALKPGVYTKVFDYNDWIQSIIAGNTAATCPP 2571GDDPCGYVLQPALYTGVFDHLDWIQSIIAGNTAATCPP 2571GDDPCGYVLQPALYTGVFDHLDWIQSIIAGNTDATCPP 25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP 2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP 24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP 2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP 273			
2CVNINLLNYSVCRAAYPEYGLPATSRTLCAGILEGGKDTCVGDSGGPLICNGQFQGIASW22111CANINILDQAVCREIYPGLLEKSRVLCAGILEGGKDTCGGDSGGPLICNGEIQGILSV2196CANINLLDYEVCRAAHPQFGLPATSRILCAGVLEGGIDTCHRDSGGPLICNGQFQGIVSW20410CANINLLDYEVCRAAYSSLPATSRVLCAGVLEGGIDTCNRDSGGPLICDQQFQGIVSW2227CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW2355GGDICAQPREPGVYTKVFDYTDWIQNIIAGNTDATCPP2589GPDTCAQPDKPSLYTKVFDHLDWIQSIIAGNTDATCPP2588GAHPCGRRLNPGFYTKVFDYIDWIQSIIAGNTDATCPP2583GAHPCGQALKPGVYTKVFDYNDWIQSIIAGNTAATCPP2571GDDPCGYVLQPALYTGVFDHLDWIQSIIAGNTDATCPP25911GGDPCAQSHVPALYIKVFDYLWIQSIIAGNTDATCPP25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQSIIAGNTDATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQSIIAGNTDATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQSIIAGNTDATCPP2577GADTCAQPRAPGFYTNVFDYTEWIQSIITGNTAATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQSIIAGNTDATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQSIIAGNTDATCPP2577GADTCAQPRAPGFYTNVFDYTEWIQSIITGNTAATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP273			
11CANINILDQAVCREIYPGLLEKSRVLCAGILEGGKDTCGGDSGGPLICNGEIQGILSV 2196CANINLLDYEVCRAAHPQFGLPATSRILCAGVLEGGIDTCHRDSGGPLICNGQFQGIVSW 20410CANINLLDYEVCRAAYSSLPATSRVLCAGVLEGGIDTCNRDSGGPLICDGQFQGIVSW 2227CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW 2355GGDICAQPREPGVYTKVFDYTDWIQNIIAGNTDATCPP 2589GPDTCAQPDKPSLYTKVFDHLDWIQSIIAGNTDATCPP 2588GAHPCGQVLTPGFYTTVFDYIDWIQSIIAGNTDATCPP 2588GAHPCGRRLNPGFYTKVFDYIDWIQSIIAGNTDATCPP 2583GAHPCGQALKPGVYTKVFDYIDWIQSIIAGNTDATCPP 2571CDDPCGYVLQPALYTGVFDHLDWIQSIIAGNTDATCPP 2572GSPNCGYVREPGLYTKVFDYNDWIQSIIAGNTDATCPP 25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP 2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP 24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP 2607GADTCAQPRAPGFYTNVFDYTDWIQNIIAGNTTVNCPP 273			
6CANINLLDYEVCRAAHPQFGLPATSRILCAGVLEGGIDTCHRDSGGPLICNGQFQGIVSW 20410CANINLLDYEVCRAAYSSLPATSRVLCAGVLEGGIDTCNRDSGGPLICDGQFQGIVSW 2227CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW 2355GGDICAQPREPGVYTKVFDYTDWIQNIIAGNTDATCPP 2589GPDTCAQPDKPSLYTKVFDYIDWIQSIIAGNKTVNCPP 2604GEHPCGQVLTPGFYTTVFDYIDWIQSIIAGNTDATCPP 2588GAHPCGRRLNPGFYTKVFDYIDWIQSIIAGNTDATCPP 2583GAHPCGQALKPGVYTKVFDYNDWIQSIIAGNTDATCPP 2583GAHPCGQALKPGVYTKVFDYNDWIQSIIAGNTDATCPP 2571GDDPCGYVLQPALYTGVFDHLDWIQSIIAGNTDATCPP 2622GSPNCGYVREPGLYTKVFDHLDWIQSIIAGNTDATCPP 25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP 2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP 24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP 2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP 273			
10CANINLLDYEVCRAAYSSLPATSRVLCAGVLEGGIDTCNRDSGGPLICDGQFQGIVSW 2227CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW 2355GGDICAQPREPGVYTKVFDYTDWIQNIIAGNTDATCPP 2589GPDTCAQPDKPSLYTKVFDHLDWIQSIIAGNKTVNCPP 2604GEHPCGQVLTPGFYTTVFDYIDWIQSIIAGNTDATCPP 2588GAHPCGRRLNPGFYTKVFDYIDWIQSIIAGNTDATCPP 2583GAHPCGQALKPGVYTKVFDYIDWIQSIIAGNTATCPP 2571GDDPCGYVLQPALYTGVFDHLDWIQSIIAGNTDATCPP 2622GSPNCGYVREPGLYTKVFDYLDWIQSIIAGNTDATCPP 25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP 2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP 24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP 2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP 273			
7CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW 2355GGDICAQPREPGVYTKVFDYTDWIQNIIAGNTDATCPP 2589GPDTCAQPDKPSLYTKVFDHLDWIQSIIAGNKTVNCPP 2604GEHPCGQVLTPGFYTTVFDYIDWIQSIIAGNTDATCPP 2588GAHPCGRRLNPGFYTKVFDYIDWIQSIIAGNTATCPP 2583GAHPCGQALKPGVYTKVFDYIDWIQSIIAGNTAATCPP 2571GDDPCGYVLQPALYTGVFDHLDWIQSIIAGNTDATCPP 2622GSPNCGYVREPGLYTKVFDHLDWIQSIIAGNTDATCPP 25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP 2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP 24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP 2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP 273	6	CANINLLDYEVCRAAHPQFGLPATSRILCAGVLEGGIDTCHRDSGGPLICNGQFQGIVSW	204
5GGDICAQPREPGVYTKVFDYTDWIQNIIAGNTDATCPP2589GPDTCAQPDKPSLYTKVFDHLDWIQSIIAGNKTVNCPP2604GEHPCGQVLTPGFYTTVFDYIDWIQSIIAGNTDATCPP2588GAHPCGRRLNPGFYTKVFDYIDWIQSITAGNTTVTCPQ2583GAHPCGQALKPGVYTKVFDYNDWIQSIIAGNTAATCPP2571GDDPCGYVLQPALYTGVFDHLDWIQSIIAGNTDATCPP2622GSPNCGYVREPGLYTKVFDHLDWIQSIIAGNTDATCPP25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP273	10	CANINLLDYEVCRAAYSSLPATSRVLCAGVLEGGIDTCNRDSGGPLICDGQFQGIVSW	222
9GPDTCAQPDKPSLYTKVFDHLDWIQSIIAGNKTVNCPP2604GEHPCGQVLTPGFYTTVFDYIDWIQSIIAGNTDATCPP2588GAHPCGRRLNPGFYTKVFDYIDWIQSITAGNTTVTCPQ2583GAHPCGQALKPGVYTKVFDYNDWIQSIIAGNTAATCPP2571GDDPCGYVLQPALYTGVFDHLDWIQSIIAGNTDATCPP2622GSPNCGYVREPGLYTKVFDYLDWIQSIIAGNTDATCPP25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP273	7	CANINILDYEVCQAAYGGSPATSRTLCAGILEGGKDSCDGDSGGPLICNGQFQGIVSW	235
9GPDTCAQPDKPSLYTKVFDHLDWIQSIIAGNKTVNCPP2604GEHPCGQVLTPGFYTTVFDYIDWIQSIIAGNTDATCPP2588GAHPCGRRLNPGFYTKVFDYIDWIQSITAGNTTVTCPQ2583GAHPCGQALKPGVYTKVFDYNDWIQSIIAGNTAATCPP2571GDDPCGYVLQPALYTGVFDHLDWIQSIIAGNTDATCPP2622GSPNCGYVREPGLYTKVFDYLDWIQSIIAGNTDATCPP25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP273			
9GPDTCAQPDKPSLYTKVFDHLDWIQSIIAGNKTVNCPP2604GEHPCGQVLTPGFYTTVFDYIDWIQSIIAGNTDATCPP2588GAHPCGRRLNPGFYTKVFDYIDWIQSITAGNTTVTCPQ2583GAHPCGQALKPGVYTKVFDYNDWIQSIIAGNTAATCPP2571GDDPCGYVLQPALYTGVFDHLDWIQSIIAGNTDATCPP2622GSPNCGYVREPGLYTKVFDYLDWIQSIIAGNTDATCPP25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP273	F		
4GEHPCGQVLTPGFYTTVFDYIDWIQSIIAGNTDATCPP2588GAHPCGRRLNPGFYTKVFDYIDWIQSITAGNTTVTCPQ2583GAHPCGQALKPGVYTKVFDYNDWIQSIIAGNTAATCPP2571GDDPCGYVLQPALYTGVFDHLDWIQSIIAGNTDATCPP2622GSPNCGYVREPGLYTKVFDHLDWIQSIIAGNTDATCPP25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP273			
8GAHPCGRRLNPGFYTKVFDYIDWIQSITAGNTTVTCPQ2583GAHPCGQALKPGVYTKVFDYNDWIQSIIAGNTAATCPP2571GDDPCGYVLQPALYTGVFDHLDWIQSIIAGNTDATCPP2622GSPNCGYVREPGLYTKVFDHLDWIQSIIAGNTDATCPP25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP273		-	
3GAHPCGQALKPGVYTKVFDYNDWIQSIIAGNTAATCPP2571GDDPCGYVLQPALYTGVFDHLDWIQSIIAGNTDATCPP2622GSPNCGYVREPGLYTKVFDHLDWIQSIIAGNTDATCPP25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP273			
1GDDPCGYVLQPALYTGVFDHLDWIQSIIAGNTDATCPP2622GSPNCGYVREPGLYTKVFDHLDWIQSIIAGNTDATCPP25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP273	8	GAHPCGRRLNPGFYTKVFDYIDWIQSITAGNTTVTCPQ 258	
2GSPNCGYVREPGLYTKVFDHLDWIQSIIAGNTDATCPP25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP273	3	GAHPCGQALKPGVYTKVFDYNDWIQSIIAGNTAATCPP 257	
2GSPNCGYVREPGLYTKVFDHLDWIQSIIAGNTDATCPP25911GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP273		GDDPCGYVLQPALYTGVFDHLDWIQSIIAGNTDATCPP 262	
11GGDPCAQSHVPALYIKVFDYTEWIQSIITGNTAATCPP2576GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP273			
6GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP24210GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP273			
10GDGSCAQPDKPALYSKVFDHLDWIQNIIAGNTTVNCPP2607GADTCAQPRAPGFYTNVFDYTDWIQSIISGNTDATCPP273			
7 GADTCAQPR <i>APGFYTNVFDYTDWIQS</i> IISGNTDATCPP 273			
· · · · · · · · · · · · · · · · · · ·			
Fig 3. Sequence alignment of the serine proteinases of the venom of S. c. edwardsii. Different colors indicate the unique peptid	/	GADTCAQFKAFGFTTNVFDITDW1Q5115GNTDATCPP Z/3	
	Fig 3. Sequence align	ment of the serine proteinases of the venom of S. c. edwardsii. Different colors indicate the u	nique peptic

Fig 3. Sequence alignment of the serine proteinases of the venom of S. c. edwardsii. Different colors indicate the unique peptides identified by tandem mass spectrometry of the corresponding proteins.

doi:10.1371/journal.pone.0092091.g003

lectins in the venom of *S. c. edwardsii* to a total of nine isoforms and might indicate a prominent role of this protein family to the envenomation of prey by *S. c. edwardsii*.

Snake venom L-amino acid oxidases (SV-LAAOs) catalyze the oxidative deamidation of amino acids and, besides effects on platelet aggregation, may induce apoptosis in prey [<u>36</u>]. We identified the predicted SV-LAAO (65% coverage) from *S. c. edwardsii* and also found five additional SV-LAAOs with sequence identities related to viperid and elapid snakes; an investigation of the biological functions of these isoforms in the venom could illuminate a broader role for SV-LAAOs in envenomation.

Bradykinin-potentiating peptides (BPPs) inhibit the activity of angiotensin I-converting enzyme (ACE) by repressing both the generation of the hypertensive peptide angiotensin II as well as the degradation of the hypotensive peptide bradykinin [37]. The result of these synergistic actions is a significantly reduced blood pressure in envenomated animals [38]. The transcriptome investigation of *S. c. edwardsii* revealed only one singleton (transcript abundance 0.28%) encoding for BPPs. This low abundance is in line with the modest sequence coverage (16.8%, Fig 1) [8] obtained via proteome analysis and it appears that contrary to other pitvipers such as *Bothrops* and *Lachesis*, BPPs play a minor role in envenomation by *S c edwardsii*.

Vascular endothelial growth factors from snake venoms (VEGF-F) bind specifically the kinase insert domain-containing receptor (KDR) and thereby induce low blood pressure as well as proliferation of vascular endothelial cells [<u>39</u>, <u>40</u>]. In the venom of *S. c. edwardsii* we identified the predicted VEGF-F together with two peptides that showed homology to VEGF-Fs from the viper *B. asper* (Central America) and the elapid *C. nigrescens* (eastern small-eyed snake; coastal eastern states of Australia). Again, there appears to be greater diversity in the proteome of *S. c. edwardsii* than was previously observed.

Several protein families were identified in the current study which were not found in the previous transcriptome analysis of the venom gland of *S. c. edwardsii*. The cyclization of N-terminal glutamine by glutaminyl cyclase (QC) is an important posttranslational process in the modification of a variety of proteins including hormones and cytokines [41], and this modification is found in many venom proteins, including SVMPs and some colubrid three-finger toxins [42]. The formation of pyroglutamine at the N-terminal likely protects proteins from enzymatic degradation and induces conformational changes to improve receptor binding [43]. Recently, glutaminyl cyclase was found in the venom gland of colubrid snakes (*Boiga*) and the authors propose that this modification might lead to increased stability of venom components against exopeptidase degradation and therefore indirectly contributing to venom toxicity [44]. The QC encountered in the venom of *S. c. edwardsii* might have similar functions (Table 2). Interesting-ly, recent proteomic studies of the venoms of rattlesnakes of the *Crotalus species*, which is related to *S. c. edwardsii*, also revealed the presence of glutaminyl cyclase [45–47].

We also found in the venom proteome three peptides that showed identity to ecto-5'-nucleotidase (5' NT) from *Gloydius blomhoffi* venom. Nucleotidases from different snake venoms have been functionally related to the inhibition of platelet aggregation [48, 49]. A study using mouse and human blood revealed that 5' NT from *Crotalus atrox* inhibits platelet aggregation via the production of increased levels of extracellular adenosine [50].

Renin-like aspartic protease was described for the first time in the venom gland transcriptome of *Echis ocellatus*, a viperid snake found in West Africa [51]. Based on the confident identification of a single peptide we confirm the presence of this enzyme in the proteome of the venom of *S. edwardsii*. To date, there are no further descriptions in the literature on the potential function of this protein, and it would be interesting to investigate the biological implications of this enzyme, especially on venom potency.

Phospholipase B (PLB) cleaves ester linkages from both the sn-1 and sn-2 positions of glycerophospholipids. Recently, a PLB has been identified for the first time in the venom of the



Protein family	Taxonomy	gi	Sequence	charge	m/z	score
Ecto-5'-nucleotidase	Gloydius blomhoffi	211926756	SSGNPILLNK	2	521.80	73.51
			ETPVLSNPGPYLEFR	1	1718.83	95
			LTAVLPFGGTFDLLQIK	1	1834.08	0.58
Glutaminyl cyclase	Gloydius blomhoffi	15991080	LIFFDGEEAFVR	2	721.88	75.11
			TFSNIISTLNPLAK	2	759.94	69.35
			WSPSDSLYGSR	2	627.8	54.71
			FVLLDLIGAR	2	558.85	52.56
			NTYQIQGIDLFVLLDLIGAR	1	2263.29	0.53
Renin-like aspartic protease	Echis ocellatus	109287598	GFLSQDIVR	1	1034.57	0.49
Phospholipase B	Crotalus adamanteus	338855308	VVPESLFAWER	1	1332.72	74
			HGLEFSYEMAPR	1	1436.66	97
			NGYWPSYNIPFDK	1	1600.77	67
			HQGLPESYNFDFVTMKPVL	1	2222.07	48

Peptides scores of the different search engines are: Peaks-regular, Mascot-bold, and ProLuCID-italics letters.

doi:10.1371/journal.pone.0092091.t002

cryptic snake *Drysdalia coronoides* [21]. While the three dimensional structure of this enzyme have not yet been resolved, it is known to form both monomers and dimers. The peptides of the PLB encountered in the venom of S. c. edwardsii show identity to the PLB detected in the transcriptome of the venom gland of the related rattlesnake C. adamanteus. The entire protein sequence of the PLB in the venom of C. adamanteus shows 553 amino acids, including a 27 residue signal peptide and a 526 amino acid phospholipase B domain [52]. Elucidation of the structure and function(s) of this protein from S. c. edwardsii venom may reveal diverse phospholipase subtypes in this venom and help explain the lack of PLA₂ diversity. It is interesting to compare the sequences identified in the venom of S. c. edwardsii in the present study with the sequences detected by Edman degradation and *de novo* sequencing of MS/MS spectra in the same venom in the study by Sanz and coworkers. While we were able to match nearly all of the sequences determined by N-terminal sequencing (S3 Table) in the study by Sanz et al. [7] to protein snake venom families of S. c. edwardsii identified in the present work, peptide sequences inferred by *de novo* sequencing (S3 Table) were more difficult to match. Interestingly, many sequences in the paper by Sanz et al. refer to SVMP's (S2 Table), which we were not able identify in the venom of S. c. edwardsii, in spite of the fact that both studies utilized venom from the same source population. This might point to venom heterogeneity among this population of snakes occurring in a rather limited area (~1600 hectares).

The relatively high abundance of metalloproteinases in the venom of *S. edwardsii*, which have the ability to cleave extracellular matrix and other structural proteins, indicate that the envenomation of prey is primarily related to hemorrhagic/tissue damaging events rather than myotoxic effects. This conclusion is also supported by the observation that specific small peptide myotoxins, such as myotoxin a from *C. viridis viridis* venom [53], and prominent PLA₂ myotoxins [54], appear to be absent from the venom. Human envenomations by *Sistrurus catenatus* are uncommon; for example, only 9/650 reported snakebites resulted from *S. catenatus* [55]. Similarly, case reports are rare, but the clinical presentation is considered to be similar to *Crotalus* sp. bites, requiring antivenom treatment but typically with less severe outcome [56, 57]. Bites by *S. c. edwardsii* are even less frequent, but because this species has a toxic venom (mouse $LD_{50} = 0.60 \ \mu g/g; [58]$) which contains abundant serine proteases, coagulopathies including hypofibrinogenemia and thrombocytopenia are to be expected.

Proteome of the venom of S. c. edwardsii

The sequence coverages accomplished by mass spectrometry of the different venom proteins range from 16.8% (BPP) to 82% (PLA₂). This distribution is reflected in the corresponding transcriptome analysis of the venom of S. c. edwardsii, in which the BPP represents the lowest abundance protein (0.28%) and the PLA₂ the highest abundance protein (28.06%) of the single protein identifications. The SVMPs of the venom of S. c. edwardsii reveal moderate sequence coverage of up 56% that might be related to the fact that the prodomain of the corresponding isoforms is included in the sequence of the mature proteins. However, the prodomain of SVMPs might be proteolytically cleaved *before* secretion into the venom gland lumen [16]. In this case, the sequence of the mature SVMPs would lack the prodomain, hence, the coverage would significantly increase. The differentiation of multiple isoforms of proteins by mass spectrometry is particularly demanding due to pronounced sequence homology, as is the case with SVSPs from the venom of S. c. edwardsii. However, we were able to distinguish ten out of eleven predicted isoforms of SVSPs (Fig 2) based on the identification of unique peptides of the corresponding proteins. The complementary use of ESI and MALDI ionization techniques leads to increased sequence coverage of the proteins investigated, compared to the sole application of one of these techniques [59, 60]. Indeed, inspection of the peptides detected revealed the identification of different sets of peptides of the corresponding proteins depending on the ionization technique applied. In some cases, such as the C-type lectin isoforms and the vascular endothelial growth factor, peptides were predominantly identified by MALDI, while other proteins like CRISP and the L-amino acid oxidase were detected by the identification of tryptic peptides ionized primarily by ESI. Different studies have shown that in MALDI experiments, peptides containing arginine residues generated through tryptic digestion are preferentially ionized compared to peptides that carry a lysine residue [61-63]. This has been related to the increased gas phase basicity of arginine compared to lysine. During the course of the manual MS/MS spectra analysis we also noted that rather large peptides (>2500 Da) revealed in many cases improved sequence fragmentation when detected by MALDI-TOF/TOF, compared to the same peptide analyzed by ESI in the linear trap. The use of ETD (electron transfer dissociation) [64] as a fragmentation technique of the tryptic peptides had only a minor impact on the improvement of the sequence coverage of the proteins investigated. However, triply charged precursor ions fragmented by ETD yielded (in some cases) more complete series of product ions and therefore more extensive sequence information when compared with the corresponding CID spectra. It is also important to note that the database search of the MS/MS spectra revealed substantially more positive results when semi-tryptic sequences were considered compared to the fully tryptic approach (dual tryptic termini). However, database searches including no specification of the enzyme did not improve protein identifications. This could possibly be explained as a significant increase in the search space, which ultimately reduces the search engine's sensitivity [65].

Conclusions

The combined approach of electrospray and MALDI mass spectrometry increased the sequence coverage of the predicted protein families (metalloproteinases, serine-proteinases, CRISP, C-type lectin, L-amino acid oxidase, vascular endothelial growth factor, bradykinin-potentiating protein and phospholipase A₂) and their corresponding isoforms when compared to one ionization technique alone. Additionally, this approach also revealed the presence of snake venom protein families (glutaminyl cyclase, renin-like aspartic protease and ecto 5'-nucleotidase) previously not encountered in the transcriptome of the venom gland of *S.c edwardsii*. These results support the use of a dual technical approach toward determining the proteome of venoms, which have both abundant and rare protein components, in order to obtain a more complete analysis. Our results revealed increased diversity of venom constituents in this venom and provide support for future studies of structure-function relationships of several venom protein family isoforms.

Supporting Information

S1 Table. Sequences of venom proteins from *S. c. edwardsii* identified by tandem mass spectrometry. Search engine color code for the identified peptides: Peaks white, ProLucid grey, and Mascot light blue.

(DOC)

S2 Table. Comparison of the sequences determined by *de novo* sequencing of MS/MS spectra of the venom of *S. edwardsii* (Sanz *et al.*) with the sequences detected in the present study.

(DOC)

S3 Table. Comparison of the sequences determined by Edmann sequencing of the venom of *S. edwardsii* (Sanz *et al.*) with the sequences detected in the present study. (DOC)

S4 Table. Sequence coverages of some of the predicted venom gland proteins of *S. c. edwardsii* as revealed by the combined approach of MALDI and ESI tandem mass spectrometry. Protein sequences including specific domains are indicted by colored bars; below these, corresponding peptides identified by ESI (black lines) and MALDI (red lines) are indicated (Part 1).

(DOCX)

S5 Table. Sequence coverages of some of the predicted venom gland proteins of *S. c. edwardsii* as revealed by the combined approach of MALDI and ESI tandem mass spectrometry. Protein sequences including specific domains are indicted by colored bars; below these, corresponding peptides identified by ESI (black lines) and MALDI (red lines) are indicated (Part 2).

(DOCX)

Author Contributions

Conceived and designed the experiments: AC RMK JP. Performed the experiments: AC ATF QL MAR. Analyzed the data: AC PCC RV. Contributed reagents/materials/analysis tools: AC JP PCC RMK. Wrote the paper: AC SPM.

References

- Lukas RJ, Morimoto H, Hanley MR, Bennett EL (1981) Radiolabeled alpha-bungarotoxin derivatives: kinetic interaction with nicotinic acetylcholine receptors. Biochemistry 20: 7373–7378. PMID: <u>7326232</u>
- OmPraba G, Chapeaurouge A, Doley R, Devi KR, Padmanaban P, et al. (2010) Identification of a novel family of snake venom proteins Veficolins from *Cerberus rynchops* using a venom gland transcriptomics and proteomics approach. J Proteome Res 9: 1882–1893. doi: <u>10.1021/pr901044x</u> PMID: <u>20158271</u>
- Wagstaff SC, Sanz L, Juarez P, Harrison RA, Calvete JJ (2009) Combined snake venomics and venom gland transcriptomic analysis of the ocellated carpet viper, *Echis ocellatus*. J Proteomics 71: 609–623. doi: <u>10.1016/j.jprot.2008.10.003</u> PMID: <u>19026773</u>
- 4. Calvete JJ, Ghezellou P, Paiva O, Matainaho T, Ghassempour A, et al. (2012) Snake venomics of two poorly known *Hydrophiinae*: Comparative proteomics of the venoms of terrestrial *Toxicocalamus*

longissimus and marine *Hydrophis cyanocinctus*. J Proteomics 75: 4091–4101. doi: <u>10.1016/j.jprot.</u> <u>2012.05.026</u> PMID: <u>22643073</u>

- 5. Holycross AT, Mackessy SP (2002) Variation in the diet of *Sistrurus catenatus* (Massasauga), with emphasis on *Sistrurus catenatus edwardsii* (Desert Massasauga). J Herpetol 36: 454–464.
- 6. Hobert JP, Montgomery CE, Mackessy SP (2004) Natural history of the Massasauga Sistrurus catenatus edwardsii, in Southeastern Colorado. Southwest Nat 49: 321–326.
- Sanz L, Gibbs HL, Mackessy SP, Calvete JJ (2006) Venom proteomes of closely related Sistrurus rattlesnakes with divergent diets. J Proteome Res 5: 2098–2112. PMID: <u>16944921</u>
- Pahari S, Mackessy SP, Kini RM (2007) The venom gland transcriptome of the Desert Massasauga rattlesnake (Sistrurus catenatus edwardsii): towards an understanding of venom composition among advanced snakes (Superfamily Colubroidea). BMC Mol Biol 8: 115–132. PMID: <u>18096037</u>
- 9. Mackessy SP (1988) Venom ontogeny in the pacific rattlesnakes *Crotalus-viridis-helleri* and *Crotalus-viridis-oreganus*. Copeia: 92–101.
- 10. Xu T, Venable JD, Park SK, Cociorva D, Lu B, et al. (2006) ProLuCID, a fast and sensitive tandem mass spectra-based protein identification program. Mol Cell Proteomics 5: S174.
- Carvalho PC, Fischer JSG, Xu T, Cociorva D, Balbuena TS, et al. (2012) Search engine processor: Filtering and organizing peptide spectrum matches. Proteomics 12: 944–949. doi: <u>10.1002/pmic.</u> 201100529 PMID: 22311825
- Gutierrez JM, Rucavado A, Escalante T, Diaz C (2005) Hemorrhage induced by snake venom metalloproteinases: biochemical and biophysical mechanisms involved in microvessel damage. Toxicon 45: 997–1011. PMID: 15922771
- Gutierrez JM, Rucavado A (2000) Snake venom metalloproteinases: Their role in the pathogenesis of local tissue damage. Biochimie 82: 841–850. PMID: 11086214
- Teixeira CF, Fernandes CM, Zuliani JP, Zamuner SF (2005) Inflammatory effects of snake venom metalloproteinases. Mem Inst Oswaldo Cruz 100: 181–184. PMID: <u>15962120</u>
- Lu Q, Navdaev A, Clemetson JM, Clemetson KJ (2005) Snake venom C-type lectins interacting with platelet receptors. Structure-function relationships and effects on haemostasis. Toxicon 45: 1089– 1098. PMID: 15876445
- Fox JW, Serrano SM (2008) Insights into and speculations about snake venom metalloproteinase (SVMP) synthesis, folding and disulfide bond formation and their contribution to venom complexity. Febs J 275: 3016–3030. doi: 10.1111/j.1742-4658.2008.06466.x PMID: 18479462
- Fox JW, Serrano SMT (2009) Timeline of key events in snake venom metalloproteinase research. J Proteomics 72: 200–209. doi: 10.1016/j.jprot.2009.01.015 PMID: 19344655
- Takeda S, Takeya H, Iwanaga S (2012) Snake venom metalloproteinases: structure, function and relevance to the mammalian ADAM/ADAMTS family proteins. Biochim Biophys Acta 1824: 164–176. doi: 10.1016/j.bbapap.2011.04.009 PMID: 21530690
- Cominetti MR, Ribeiro JU, Fox JW, Selistre-de-Araujo HS (2003) BaG, a new dimeric metalloproteinase/disintegrin from the *Bothrops alternatus* snake venom that interacts with alpha5beta1 integrin. Arch Biochem Biophys 416: 171–179. PMID: 12893294
- Valente RH, Guimaraes PR, Junqueira M, Neves-Ferreira AGC, Soares MR, et al. (2009) Bothrops insularis venomics: A proteomic analysis supported by transcriptomic-generated sequence data. J Proteomics 72: 241–255. doi: 10.1016/j.jprot.2009.01.001 PMID: 19211044
- Chatrath ST, Chapeaurouge A, Lin Q, Lim TK, Dunstan N, et al. (2011) Identification of novel proteins from the venom of a cryptic snake *Drysdalia coronoides* by a combined transcriptomics and proteomics approach. J Proteome Res 10: 739–750. doi: <u>10.1021/pr1008916</u> PMID: <u>21133350</u>
- 22. Rokyta DR, Lemmon AR, Margres MJ, Aronow K (2012) The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). BMC Genomics 13.
- Juarez P, Comas I, Gonzalez-Candelas F, Calvete JJ (2008) Evolution of snake venom disintegrins by positive Darwinian selection. Mol Bio Evol: 2391–2407.
- 24. Markland FS (1998) Snake venoms and the hemostatic system. Toxicon 36: 1749–1800. PMID: 9839663
- 25. Pirkle H (1998) Thrombin-like enzymes from snake venoms: an updated inventory. Thromb Haemost 79: 675–683. PMID: 9531061
- Serrano SM, Maroun RC (2005) Snake venom serine proteinases: sequence homology vs. substrate specificity, a paradox to be solved. Toxicon 45: 1115–1132. PMID: 15922778
- 27. Lee JW, Seu JH, Rhee IK, Jin I, Kawamura Y, et al. (1999) Purification and characterization of brevinase, a heterogeneous two-chain fibrinolytic enzyme from the venom of Korean snake, *Agkistrodon blomhoffii brevicaudus*. Biochem Biophys Res Commun 260: 665–670. PMID: <u>10403823</u>

- 28. Kini RM (1997) In: Kini RM (ed). John Wiley & Sons: Chichester, England p 1–28
- Kini RM (2006) Anticoagulant proteins from snake venoms: structure, function and mechanism. Biochem J 397: 377–387. PMID: <u>16831131</u>
- Yamazaki Y, Morita T (2004) Structure and function of snake venom cysteine-rich secretory proteins. Toxicon 44: 227–231. PMID: <u>15302528</u>
- **31.** Shikamoto Y, Suto K, Yamazaki Y, Morita T, Mizuno H (2005) Crystal structure of a CRISP family Ca2 +-channel blocker derived from snake venom. J Mol Biol 350: 735–743. PMID: <u>15953617</u>
- Yamazaki Y, Koike H, Sugiyama Y, Motoyoshi K, Wada T, et al. (2002) Cloning and characterization of novel snake venom proteins that block smooth muscle contraction. Eur J Biochem 269: 2708–2715. PMID: <u>12047379</u>
- Yamazaki Y, Hyodo F, Morita T (2003) Wide distribution of cysteine-rich secretory proteins in snake venoms: isolation and cloning of novel snake venom cysteine-rich secretory proteins. Arch Biochem Biophys 412: 133–141. PMID: <u>12646276</u>
- Clemetson KJ, Morita T, Kini RM (2009) Classification and nomenclature of snake venom C-type lectins and related proteins. Toxicon 54: 83. doi: <u>10.1016/j.toxicon.2009.04.001</u> PMID: <u>19362105</u>
- Hirabayashi J, Kusunoki T, Kasai K (1991) Complete primary structure of a galactose-specific lectin from the venom of the rattlesnake *Crotalus atrox*. Homologies with Ca2(+)-dependent-type lectins. J Biol Chem 266: 2320–2326. PMID: 1989986
- Du XY, Clemetson KJ (2002) Snake venom L-amino acid oxidases. Toxicon 40: 659–665. PMID: 12175601
- Villard E, Soubrier F (1996) Molecular biology and genetics of the angiotensin-l-converting enzyme: potential implications in cardiovascular diseases. Cardiovasc Res 32: 999–1007. PMID: <u>9015402</u>
- Linz W, Wiemer G, Gohlke P, Unger T, Schölkens BA (1995) Contribution of kinins to the cardiovascular actions of angiotensin-converting enzyme inhibitors. Pharmacol Rev 47: 25–49. PMID: <u>7784479</u>
- Yamazaki Y, Takani K, Atoda H, Morita T (2003) Snake venom vascular endothelial growth factors (VEGFs) exhibit potent activity through their specific recognition of KDR (VEGF receptor 2). J Biol Chem 278: 51985–51988. PMID: 14600159
- 40. Yamazaki Y, Matsunaga Y, Nakano Y, Morita T (2005) Identification of vascular endothelial growth factor receptor-binding protein in the venom of eastern cottonmouth. A new role of snake venom myotoxic Lys49-phospholipase A2. J Biol Chem 280: 29989–29992. PMID: <u>16014630</u>
- Huang KF, Liu YL, Cheng WJ, Ko TP, Wang AH (2005) Crystal structures of human glutaminyl cyclase, an enzyme responsible for protein N-terminal pyroglutamate formation. Proc Natl Acad Sci U S A 102: 13117–13122. PMID: <u>16135565</u>
- Pawlak J, Mackessy SP, Sixberry NM, Stura EA, Le Du MH, et al. (2009) Irditoxin, a novel covalently linked heterodimeric three-finger toxin with high taxon-specific neurotoxicity. FASEB J 23: 534–545. doi: 10.1096/fj.08-113555 PMID: 18952712
- 43. Hinke S, Pospisilik JA, Demuth HU, Mannhart S, Kühn-Wache K, et al. (2000) Dipeptidyl peptidase IV (DPIV/CD26) degradation of glucagon. Characterization of glucagon degradation products and DPIV-resistant analogs. J Biol Chem 275: 3827–3834. PMID: 10660533
- 44. Pawlak J, Kini RM (2006) Snake venom glutaminyl cyclase. Toxicon 48: 278–286. PMID: 16863655
- 45. Georgieva D, Ohler M, Seifert J, von Bergen M, Arni RK, et al. (2010) Snake venomics of Crotalus durissus terrificus—correlation with pharmacological activities. J Proteome Res 9: 2302–2316. doi: <u>10.</u> 1021/pr901042p PMID: 20205475
- 46. Calvete JJ, Fasoli E, Sanz L, Boschetti E, Righetti PG (2009) Exploring the Venom Proteome of the Western Diamondback Rattlesnake, Crotalus atrox, via Snake Venomics and Combinatorial Peptide Ligand Library Approaches. J Proteome Res 8: 3055–3067. doi: <u>10.1021/pr9002499</u> PMID: <u>19371136</u>
- Fox JW, Ma L, Nelson K, Sherman NE, Serrano SMT (2006) Comparison of indirect and direct approaches using ion-trap and Fourier transform ion cyclotron resonance mass spectrometry for exploring viperid venom proteomes. Toxicon 47: 700–714. PMID: 16574175
- Kini RM, Evans HJ (1990) Effects of snake-venom proteins on blood-platelets. Toxicon 28: 1387– 1422. PMID: <u>2089737</u>
- Ouyang C, Huang TF (1983) Inhibition of platelet-aggregation by 5'-nucleotidase purified from *Trimere-surus-gramineus* snake-venom. Toxicon 21: 491–501. PMID: <u>6312633</u>
- Hart ML, Kohler D, Eckle T, Kloor D, Stahl GL, et al. (2008) Direct treatment of mouse or human blood with soluble 5 '-nucleotidase inhibits platelet aggregation. Arterioscler Thromb Vasc Biol 28: 1477– 1483. doi: <u>10.1161/ATVBAHA.108.169219</u> PMID: <u>18511695</u>

- Wagstaff SC, Harrison RA (2006) Venom gland EST analysis of the saw-scaled viper, Echis ocellatus, reveals novel alpha9beta1 integrin-binding motifs in venom metalloproteinases and a new group of putative toxins, renin-like aspartic proteases. Gene 377: 21–32. PMID: <u>16713134</u>
- Rokyta DR, Wray KP, Lemmon AR, Lemmon EM, Caudle SB (2011) A high-throughput venom-gland transcriptome for the Eastern Diamondback Rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes. Toxicon 57: 657–671. doi: <u>10.1016/j.toxicon.2011.01.</u> 008 PMID: 21255598
- 53. Fox JW, Elzinga M, Tu AT (1979) Amino-acid sequence and disulfide bond assignment of myotoxin-a isolated from the venom of prairie rattlesnake (*Crotalus-viridis-viridis*). Biochemistry 18: 678–684. PMID: 570412
- Gutierrez JM, Ownby CL, Odell GV (1984) Skeletal-muscle regeneration after myonecrosis induced by crude venom and a myotoxin from the snake *Bothrops-asper* (Fer-de-lance). Toxicon 22: 719–731. PMID: 6523503
- 55. Russell FE (1980) Snake Venom Poisoning Scholium International, Inc, Great Neck, New York 562 pp.
- Sing K, Erickson T, Aks S, Rothenberg H, Lipscomb J (1994) Eastern Massasauga rattlesnake envenomations in an urban wilderness. J Wild Med 5: 77–87.
- 57. Barceloux DG (2008) Medical Toxicology of Natural Substances: Foods, Fungi, Medicinal Herbs, Plants, and Venomous Animals. John Wiley & Sons, Hoboken, New Jersey 1200 pp.
- 58. Gibbs HL, Mackessy SP (2009) Functional basis of a molecular adaptation: Prey-specific toxic effects of venom from Sistrurus rattlesnakes. Toxicon 53: 672–679. doi: <u>10.1016/j.toxicon.2009.01.034</u> PMID: <u>19673082</u>
- Stapels MD, Barofsky DF (2004) Complementary use of MALDI and ESI for the HPLC-MS/MS analysis of DNA-binding proteins. Anal Chem 76: 5423–5430. PMID: <u>15362902</u>
- Yang Y, Zhang S, Howe K, Wilson DB, Moser F, et al. (2007) A comparison of nLC-ESI-MS/MS and nLC-MALDI-MS/MS for GeLC-based protein identification and iTRAQ-based shotgun quantitative proteomics. J Biomol Tech 4: 226–237. PMID: 17916795
- Brancia FL, Oliver SG, Gaskell SJ (2000) Improved matrix-assisted laser desorption/ionization mass spectrometric analysis of tryptic hydrolysates of proteins following guanidination of lysine containing peptides. Rapid Commun Mass Spectrom 14: 2070–2073. PMID: <u>11085420</u>
- Beardsley RL, Karty JA, Reilly JP (2000) Enhancing the intensities of lysine-terminated tryptic peptide ions in matrix-assisted laser desorption/ionization mass spectrometry. Rapid Commun Mass Spectrom 14: 2147–2153. PMID: <u>11114023</u>
- Carabetta VJ, Li T, Shakya A, Greco TM, Cristea IM (2010) Integrating Lys-N Proteolysis and N-Terminal Guanidination for Improved Fragmentation and Relative Quantification of Singly-Charged Ions. J Am Soc Mass Spectrom 21: 1050–1060. doi: <u>10.1016/j.jasms.2010.02.004</u> PMID: <u>20207164</u>
- Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci USA 101: 9528–9533. PMID: 15210983
- Borges D, Perez-Riverol Y, Nogueira FCS, Domont GB, Noda J, et al. (2013) Effectively addressing complex proteomic search spaces with peptide spectrum matching. Bioinformatics 29: 1343–1344. doi: <u>10.1093/bioinformatics/btt106</u> PMID: <u>23446294</u>