Comparative analyses of venom-associated genes from an Old World viper, *Daboia russelii*

Neeraja M Krishnan[1] and Binay Panda[1]*

[1]Ganit Labs, Bio-IT Centre, Institute of Bioinformatics and Applied Biotechnology, Biotech Park, Electronic City Phase I, Bangalore 560100

* Corresponding author: binay@ganitlabs.in

Abstract

Molecular basis of toxin gene diversity among snakes is poorly understood. Lack of whole genome sequence information for most snakes makes studies on toxin genes and their orthologous counterparts difficult. One of the challenges in studying snake genomes is the acquisition of biological material from live animals, especially from the venomous ones. Additionally, in certain geographies, Government permission is required to handle live snakes making the process cumbersome and time-consuming. Here, we report comparative sequence analyses of toxin genes from Russell's viper (*Daboia russelii*) using whole-genome sequencing data obtained from the skin exuviate. In addition to the comparative analyses of 46 toxin-associated proteins, we present evidence of unique sequence motifs in five key toxin-associated protein domains; nerve growth factor (NGF), platelet derived growth factor (PDGF), Kunitz/Bovine pancreatic trypsin inhibitor (Kunitz BPTI), cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins (CAP) and cysteine-rich secretory protein (CRISP). We compared the venom-associated domains from Russell's viper with those from both venomous and non-venomous vertebrates and invertebrates. The *in silico* study on structures identified V11 and T35 in the NGF domain; F23 and A29 in the PDGF domain; N69, K2 and A5 in the CAP domain; and Q17 in the CRISP domain to be responsible for differences in the largest pockets across the protein domain structures in New World vipers, Old World vipers and elapids. Similarly, residues F10, Y11 and E20 appear to play an important role in the protein structures across the kunitz protein domain of viperids and

elapids. Our study sheds light on the uniqueness of these key toxin-associated proteins and their evolution in vipers.

Data deposition: Russell's viper sequence data is deposited in the NCBI SRA database under the accession number SRR5506741 and the GenBank accession numbers for the individual venom-associated genes is provided in Table S1.

Keywords: Toxin-associated genes, Russell's viper, Old World vipers, New World vipers, elapids

Introduction

Snake venom genes and their products offer an excellent model system to study gene duplication, evolution of regulatory DNA sequences, and biochemical diversity and novelty of venom proteins. Additionally, snake venoms have tremendous potential in developing new drugs and bioactive compounds (Vonk et al. 2011). Previous studies have highlighted the importance of gene duplications and/or sub-functionalization (Malhotra et al. 2010; Rokyta et al. 2011; Hargreaves et al. 2014) and transcriptional/post-transcriptional mechanisms (Casewell et al. 2014) contributing towards snake venom diversity. Venom studies so far have extensively used peptides/protein data alongside individual gene sequences or sequences of particular family members to study variations on gene structure and sequence composition. Till date, whole genome sequences of seven snake species, king cobra *Ophiophagus hannah* (Vonk et al. 2013); Burmese

61 python *Python bivitattus* (Castoe et al. 2013); rattlesnake *Crotalus*
62 *atrox* (Dowell et al. 2016), Florida pygmy
63 rattlesnake *Sistrurus miliarius barbouri* (Vicoso et al. 2013); garter
64 snake *Thamnophis elegans* (Vicoso et al. 2013); five-pacer
65 viper *Deinagkistrodon acutus* (Yin et al. 2016); and corn
66 snake *Pantherophis guttatus* (Ullate-Agote et al. 2014), have either been
67 published or their sequence information available in the public domain. In
68 addition, genome-sequencing efforts are either underway or the sequences of
69 venom-associated genes have been deposited in the databases for few
70 others (Kerkkamp et al. 2016). Out of the sequenced genomes, only a few
71 have been annotated or the annotations have been made public, a key
72 requirement for comparative analysis of genes. This, along with the lack of
73 availability of whole genome sequences and/or complete coding sequences
74 for most snakes has made studies on toxin gene orthologies and gene
75 variation among venomous snakes limiting.
76 Four snakes, Russell's viper (*Daboia russelii),* saw-scaled viper
77 (*Echis carinatus*), spectacled cobra (*Naja naja*), and common krait
78 (*Bungarus caeruleus*) are responsible for most snakebite-related mortality in
79 India (Mohapatra et al. 2011; Whitaker 2015). Russell's viper is a Old World
80 viper, member of the taxon Viperidae and subfamily Viperinae and is
81 responsible for large numbers of snakebite incidents and deaths in India. Very
82 little is known about the diversity of venom-associated genes from any viper,
83 including the only Old World viper where complete genome sequence
84 information is available (European adder, *Vipera berus berus*,
85 https://www.ncbi.nlm.nih.gov/bioproject/170536). Lack of any complete

genome annotation from this viper and other snake species reduces the scope of a detailed comparative study on toxin-associated genes. Such a study involving various groups of venomous and non-venomous snakes, in addition to other venomous vertebrates and invertebrates, will facilitate our understanding on the evolution of these genes, their diversity, and function.

One of the challenges in studying the genomes of venomous animals is related to sample acquisition. Additionally, in India, Government permission is required to catch snakes and extract blood samples from them (all snakes are protected in India under the Indian Wildlife Protection Act, 1972). Even with permission, there is a chance to adversely affect the animals during sample acquisition. This may be circumvented by the use of skin exuviate (shed skin) that does not require handling or drawing blood or taking any tissue from the animals. However, working with DNA isolated from shed skin has its own challenges. Microbial contamination, lack of full-length DNA in the exuviate cells and computational challenges in dealing with short stretches of DNA are some of the bottlenecks for working with DNA from exuviate skins.

In the current study, we explored the possibility of getting toxin gene information from low-coverage whole-genome sequencing data using skin exuviate from an Old World viper, Russell's viper, and performed comparative analysis on the annotated 51 venom-associated genes representing all the major venom-associated protein families (Fry 2005) for which coding sequences were available from a New World viper, a pit viper, *Protobothrops mucrosquamatus*. We focused our analyses on five key toxin-associated protein domains; nerve growth factor (NGF), platelet derived growth factor (PDGF), Kunitz/Bovine pancreatic trypsin inhibitor (Kunitz BPTI),

111 cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1
112 proteins (CAP) and cysteine-rich secretory protein (CRISP) in Russell's viper
113 and discovered key residues that are changed across Old World vipers, New
114 World vipers and elapids that might have contributed towards the evolution of
115 venom in vipers.
116
117 **Materials and Methods**
118 *Russell's viper skin exuviate and DNA isolation*
119 Freshly shed skin exuviate of Russell's viper from Bangalore, India was
120 a gift from Mr. Gerry Martin. The skin exuviate for the entire snake was
121 obtained, cleaned thoroughly with 70% ethanol and nuclease-free water 3
122 times each, dried thoroughly and frozen until the time of extraction of DNA.
123 Genomic DNA was extracted following the protocol of Fetzner (Fetzner
124 1999) with modifications.
125
126 *Sequencing, read processing and assembly*
127 Illumina paired-end read libraries were prepared following the
128 manufacturer instructions using amplification free genomic DNA library
129 preparation kit and sequenced using Illumina HiSeq2500 instrument. Archael,
130 bacterial and human sequence contamination were removed
131 from the Russell's viper sequence using DeConSeq (Schmieder and Edwards
132 2011) using curated and representative
133 genomes (https://www.ncbi.nlm.nih.gov/genome/browse/reference/). Furtherm
134 ore, the sequenced reads were post-processed to remove unpaired reads and
135 quality analysis was performed using FastQC v0.1

136 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The rd_len_cutoff
137 option was exercised during the read assembly step to trim off the low-
138 quality bases, since the per-base quality was found to drop below 28 after the
139 initial 50-70 bases of the read. The Russell's viper read libraries were
140 assembled using SOAPdenovo2 (r240) (Luo et al. 2012).

141 *Identifying toxin-associated genes, coding regions, and gene structures*

142 The DNA sequences for 51 out of 54 venom-associated genes (Fry
143 2005) from *Protobothrops mucrosquamatus* were downloaded (Table 1).
144 These were used to fish genomic scaffolds bearing highly similar sequences
145 in Russell's viper genome assembly, using a BLAST with an E-value threshold
146 of 10^{-3}. These scaffolds were then anchored to the respective coding
147 sequences from *Protobothrops mucrosquamatus* using a discontiguous
148 megaBlast, to determine the correct frame of translation and extract the
149 complete amino acid coding sequence (CDS) corresponding to Russell's
150 viper. We obtained the exon-intron structures for all the toxin-associated
151 genes in Russell's viper by aligning the CDS with gene sequences using
152 discontiguous megaBlast and plotted using the tool GSDS2.0 (Hu et al. 2015).
153 The sequences for the Russell's viper venom-associated genes and their
154 GenBank accession numbers are provided in Table S1.

155

156 *Comparative analyses of toxin-associated protein domains*

157 The amino acid sequences of all the Russell's viper's toxin-associated
158 genes were subjected to domain search using Pfam (Finn et al. 2016) (Table
159 S2). All domain sequences was then serially aligned using blastp to protein
160 sequences from 18 snake species (Table S3). Five domains (NGF, PDGF,

161 Kunitz BPTI, CAP and CRISP) from four genes (NGF, VEGF,
162 CRISP/Serotriflin, and Kunitoxin), which showed variability across different
163 snake groups and where sequence information were available beyond the
164 whole genome sequences, were used for expansive comparative analyses
165 (Table S4) using sequences from Viperids (taxid: 8689), elapids (taxid: 8602),
166 Colubrids (taxid: 8578), Boids (taxid: 8572), Acrochordids (taxid: 42164),
167 Pythonids (taxid: 34894), lizards (squamates (taxid: 8509) minus snakes
168 (taxid: 8570)), Crocodylia (taxid: 51964) and Testudines (taxid: 8459).
169
170 *3D structure prediction of the chosen domains*
171 Consensus sequences were determined from NGF, PDGF, Kunitz
172 BPTI, CAP and CRISP domain alignments using Simple Consensus Maker
173 (https://www.hiv.lanl.gov/content/sequence/CONSENSUS/SimpCon.html) for
174 New World vipers (NWV), Old World vipers (OWV) and elapids. The
175 consensus sequences were submitted to the Phyre2 (Kelley et al. 2015)
176 protein fold recognition server using the standard mode
177 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). The best 3D
178 model was further investigated by Phyre2 (Kelley et al. 2015) to analyze the
179 structural model using various open source tools.
180
181 **Results**
182 *Skin exuviate yielded fairly good quality DNA for genome sequencing and*
183 *near complete coding sequences for toxin-associated genes*
184 Genomic DNA isolated from the skin exuviate of Russell's viper was
185 fairly intact with most of the DNA in the size range of more than 5kbp range

186 (File S1). Sequenced short reads were assembled and then used to fish all
187 the 51 toxin-associated genes in Russell's viper (see Materials and Methods).
188 Next, we obtained the exon-intron structures for all the toxin-associated genes
189 in Russell's viper by aligning the CDS with gene sequences using
190 discontiguous megaBlast and plotted using the tools GSDS2.0 (Hu et al.
191 2015) (File S2). We found that the average length of exons in Russell's viper
192 toxin-associated genes is around 190 nucleotides (nt), matching well with the
193 lengths of other vertebrate exons (Gelfman, et al. 2012).
194
195 *Comparative protein domain analyses*
196 Among the toxin-associated genes, a larger pool of sequences are
197 available only for NGF, PDGF domain of VEGF, Kunitz_BPTI domain of
198 Kunitoxin, CRISP and CAP domains in CRISP and Serotriflin proteins, from
199 various snake groups (Colubridae, Boidae, Pythonidae and Acrochordidae),
200 non-snake reptilian groups (lizards, crocodiles and Testudines), venomous
201 invertebrates (wasps, spiders and scorpions) and venomous vertebrates
202 (fishes and mammals). Therefore, these domains were compared with those
203 from Russell's viper. Comparative domain analysis was performed for all
204 toxin-associated genes (File S3) across 18 snake species where sequence
205 information is available (Table S3). In the case of five domains: CAP and
206 CRISP domains of CRISP and serotriflin genes (L and AL), Kunitz BPTI
207 of kunitoxin (S), NGF (T) and PDGF of VEGFA (AP-AR) and VEGFF (AU), we
208 found that the maximum number of species aligned to their domain
209 sequences. Some protein domains, the CRISP, Kunitz BPTI, guanylate CYC,
210 PDGF of VEGFF and WAP, show long stretches of mismatches (File S3)

211 compared with Russell's viper sequence. Out of these, only NGF and PDGF
212 domains of VEGF carried amino acid changes specific to the New World
213 vipers, that were completely absent in any other group used for comparison,
214 including in lizards, crocodiles, and turtles (File S4). Specific changes in these
215 proteins and their implications are discussed below.

216 Russell's viper NGF is a single exon gene with a 745nt transcript
217 coding for a 244 amino acid protein consisting of a single NGF domain (fig.
218 1A). The NGF domain bears 28% sequence conservation across all the five
219 vertebrate phyla, namely, fishes, amphibians, reptiles, aves and mammals
220 distributed along the length of the domain (fig. 1B). Thirty-six percent out of
221 these residues are conserved across other venomous vertebrates
222 (fishes, squamates and mammals) and venomous invertebrates (scorpions
223 and wasps) (fig. 1C). Thirteen percent of Russell's viper NGF domain residues
224 are variable with respect to the domain sequence in at least one among the
225 NGF sequences in the groups of New World vipers, Old World vipers and
226 elapids (fig. 1D). Although several amino acids in the NGF domain in New
227 World vipers seem to have changed from the Russell's viper and other Old
228 World vipers, their function probably remains unchanged. For example,
229 phenylalanine (F) to isoleucine (I) at position 12 and serine (S) to Asparagine
230 (N) at position 19 between the Old World vipers to the New World vipers does
231 not change the function of the amino acids (from one hydrophobic amino acid
232 to another and from one polar amino acid to another). However, there are
233 others, for example, threonine (T) and glutamine (Q), at position 67 and 68
234 respectively in the NGF domain of the New World viper evolved only in that
235 group. One of those, a polar amino acid glutamine at position 68, is a very

236 important residue as its corresponding amino acid in any of the other snakes,
237 except in colubrids, is a hydrophobic proline. This evolution of NGF sequence
238 might have implications on its function in the venom of the New World vipers.
239 In Russell's viper, the VEGFA gene comprises five exons coding for a
240 652nt long transcript and a protein with two domains: PDGF and VEGF-C (fig.
241 2A). The PDGF domain sequence exhibits conservation in 65% of its residues
242 across the three vertebrate phyla (reptiles, aves and mammals) (fig. 2B).
243 Since sequence information from fishes and amphibians were not available,
244 they could not be included in the comparison study. Out of the conserved
245 residues, 21% of those were also conserved in venomous vertebrates
246 (squamates and mammals) and venomous invertebrates (wasps). Fifteen
247 percent of the PDGF domain residues were variable in at least one of the
248 three snake groups: New World vipers, Old World vipers and elapids (fig.2C
249 and fig. 2D). Like the NGF domain, the evolution of the PDGF domain in New
250 World vipers at certain amino acids is striking. For example, in the New World
251 vipers, the position 67 is a polar amino acid tyrosine (Y) while in all other
252 reptiles, venomous invertebrates and mammals; this is primarily a
253 hydrophobic amino acid phenylalanine (F). This might bear implications on the
254 protein's structure and function.
255 Kunitoxin in Russell's viper is a 3.1kb gene comprising two exons, with
256 a transcript length of 270nt that codes for a 44 amino acids long single Kunitz
257 BPTI domain (fig. 3A). About 29% of the protein domain residues are
258 conserved across the four vertebrate phyla (amphibians, reptiles, aves and
259 mammals) (fig. 3B). Since sequence information from the Kunitz BPTI for
260 fishes was not available, they could not be included in the comparison. Out of

261 these conserved residues, 76% are conserved in venomous vertebrates
262 (squamates and mammals) and venomous invertebrates (scorpions and
263 wasps) (fig. 3C) and 56% of the domain residues are variable in at least one
264 of three snake groups (New World vipers, Old World vipers, and elapids) (fig.
265 3D). Of the residues that are evolved in the New World vipers, the second
266 residue, a positively charged one, alanine (A) is present only in the Old World
267 vipers, which is replaced by a hydrophobic residue, proline (P), in the New
268 World vipers and elapids. Residues 14-18 are very polymorphic in the New
269 World vipers and elapids, but not so in the Old World vipers.

270 The CRISP gene in Russell's viper is a 25kb long gene, comprises of 8
271 exons coding for a 787nt transcript and two protein domains, CAP and CRISP
272 (fig. 4A). The CAP domain exhibits conservation in 7% of its residues across
273 all the five vertebrate phyla (fig. 4B). Forty-two percent of those residues are
274 conserved across venomous vertebrates (amphibians, squamates and
275 mammals) and venomous invertebrates (scorpions and wasps) (fig. 4C). In
276 addition, there are five residues conserved across all the venomous animals
277 (fig. 4C). Twenty-seven percent of the CAP domain and 15% of the CRISP
278 domain residues are variable in at least one of the three snake groups (fig.
279 4D). There are several extra residues for the CAP domain in the New World
280 vipers and elapids, but not in the Old World vipers. The conserved residues
281 comprised mostly of Cystines and to a lesser extent Asparagines (fig. 4E)
282 across venomous vertebrates (squamates and mammals) (fig. 4F). Sixty
283 percent of the CRISP domain residues are variable in at least one viperids or
284 elapid member with respect to the domain sequence of Russel's viper (fig.
285 4G).

286 Next, we explored the role of consensus domain sequences and their
287 possible role of conserved amino acids in those key toxin-associated protein
288 domains across New World vipers, Old World vipers and elapids. We
289 constructed the 3D structure models using Phyre2, followed by Phyre2
290 investigation, for further analyses on the structural model. As evident from the
291 analyses, amino acid residues 18-19 and 117 of the NGF domain reflected a
292 difference in mutation sensitivity as detected by SusPect algorithm (Yates et
293 al. 2014), especially in the elapids compared to the viperids (fig. 5). Residue
294 18 is Valine in the viperids and Isoleucine in the elapids; residue 19 is Serine
295 in the Old World vipers and Asparagine in the New World vipers; and residue
296 117 is Threonine in the elapids and Serine in the New World vipers (fig. 5A).
297 This might have implications in the structure of the protein as the largest
298 pockets detected by fpocket algorithm appear to be vastly different among the
299 New World vipers, Old World vipers and elapids for the NGF, PDGF, CAP and
300 CRISP domains (fig. 5). The pockets appear small in all cases for
301 the elapids, and largest in the case of Old World vipers (fig. 5). Minor
302 differences in clashes were observed at residues 10,11 and 20 of the Kunitz
303 domain and residue 38 of this domain showed a rotamer conflict in the case of
304 the New World vipers (fig. 5C). Similarly, residue 46 of the CAP domain and
305 residues 4 and 31 of the CRISP domain showed rotamer conflict for the
306 viperids (fig. 5D and fig. 5E). The other protein quality and functional
307 parameters were not affected across the 3D structure models for the three
308 snake groups (File S5).
309
310 **Discussion**

311 Accessibility and affordability for high-throughput sequencing
312 technologies along with the availability of sophisticated computational tools to
313 assemble, annotate and interpret genomes is playing a powerful role in
314 deciphering gene functions and their role in evolution. Multiloci gene families
315 usually encode snake toxin genes that produce gene isoforms generated by
316 the process of gene duplications (Fry 2005; Casewell et al. 2013). Russell's
317 viper (*Daboai russelli*) is a member of the taxon *Viperidae* and
318 subfamily *Viperinae* and is one among four other venomous snakes
319 responsible for the most snakebite incidents and deaths in India (Mohapatra
320 et al. 2011; Whitaker 2015). Russell's viper is a member of the group Old
321 World vipers or true vipers or pitless vipers, characterized by the lack of heat-
322 sensing pit organs (Mallow et al. 2003). Snakes have evolved two main
323 venom types, neurotoxic (elapids that include cobras and kraits) or
324 haemotoxic (viperids that include vipers) (Fry et al. 2008). Russell's viper
325 venom is haemotoxic. There is significant variation in the venom compositions
326 of snakes, even at times within the same species as in the case of Russell's
327 viper (Jayanthi and Gowda 1988). The variation in the venom composition
328 within the same species is thought to be a result of adaptation in response to
329 the difference in diets (Daltry et al. 1996; Barlow et al. 2009; Casewell et al.
330 2013). The composition of venom of Russell's viper from different regions of
331 India is different (Jayanthi and Gowda 1988), making anti-venom produced
332 using snake venoms from a single geography against all Russell's viper bites
333 across the country ineffective. Currently, efforts are underway to collect
334 venoms of Russell's viper from different regions of India in order to
335 understand the difference in their venom composition (Rom Whitaker, and

336 Gerry Martin, *personal communications*). Currently, information on toxin-
337 associated genes from any viper, especially Russell's viper is extremely
338 limiting and therefore, comparative analyses of the toxin-associated genes will
339 add value to our understanding of its venom. The only Old World viper, where
340 complete genome sequence information is available is from a European
341 adder, *Vipera berus berus* (https://www.ncbi.nlm.nih.gov/bioproject/170536).
342 Although sequence information is there for this species, the annotation is not
343 available.

344 The aim of the current study was two folds. First, as handling and
345 getting biological material from snakes in India requires Government
346 permission and specific expertise, we wanted to test whether good quality
347 whole-genome sequence information can be obtained using skin exuviate.
348 Second, we wanted to test whether we can assemble the toxin-associated
349 genes using low-coverage sequencing data from a next-generation
350 sequencing platform, and use it for comparative genomics study. On both the
351 accounts, we found the results to be satisfactory. First, although shed skin is
352 often contaminated with bacteria and other microorganisms, using freshly
353 shed skin exuviate, we successfully isolated high molecular weight genomic
354 DNA (File S1), which was subsequently used to generate genome sequencing
355 data. This helps in obtaining biological material without having to require
356 permission to capture venomous animals and also reduce stress in animals
357 during handling. We believe this is an attractive option for generating snake
358 genome data and studying molecular evolution in snakes. Second, from the
359 sequence data, we succeeded in assembling near complete CDS for 51 out of
360 54 toxin-associated genes (Fry 2005). This highlights the utility of low-

361 coverage genome sequencing data in studying important genes and their role
362 in evolution. As the lengths of the toxin-associated genes in Russell's viper
363 were much longer than the CDS, the intronic sequences were assembled with
364 gaps. This was primarily due to the low coverage sequencing data used for
365 assembly and the lack of long-insert mate pair sequencing data in our
366 repertoire. The mean length of exons for the toxin-associated genes in
367 Russell's viper is 190 base pairs, which was much smaller compared to the
368 average intron length; the exons were assembled accurately using short-read
369 sequences. Another interesting observation from our data is that the AT to GC
370 ratio in the CDS regions (cumulatively for all the 51 genes) of the toxin-
371 associated genes in Russell's viper is 1:1 whereas it is skewed (the ratio is
372 1.5:1) for the full gene sequences.

373 Our study has two implications. First, it demonstrates the feasibility
374 of *de novo* analyses of genes and gene families without prior sequence
375 information and annotation, and second; this facilitates sequencing a large
376 number of genes and gene families without going through the process of
377 designing individual primers for amplification and Sanger sequencing. Despite
378 these advantages, our study has certain limitations. First, like any other
379 annotation-based study, it relies on the quality of existing/prior annotation of
380 toxin-related genes. Although the chances are slim, as we used a closely
381 related viper to get the toxin-associated gene sequences in Russell's viper, it
382 is entirely possible that we might have missed some key genes specific to
383 Russell's viper. Using a high-coverage sequencing data along with the use
384 of *de novo* tools might help annotate unique genes in Russell's viper in the
385 future. Second, our study does not touch the expression of the toxin genes in

Russell's viper and their significance compared to other venomous snakes. It is possible that the unique spatial and temporal expression pattern of the toxin-associated genes will explain their evolution better.

Acknowledgements

This work was supported by Department of Electronics and Information Technology, Government of India [18(4)/2010-E-Infra., 31-03-2010]; and Department of IT, BT and ST, Government of Karnataka, India [3451-00-090-2-22]. We thank Kunal Dhas and Arun Hariharan for their help in laboratory experiments.

Literature Cited

Barlow, A., C. E. Pook, R. A. Harrison, and W. Wuster. 2009. Coevolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution. Proc Biol Sci 276:2443-2449.

Casewell, N. R., S. C. Wagstaff, W. Wuster, D. A. Cook, F. M. Bolton, S. I. King, D. Pla, L. Sanz, J. J. Calvete, and R. A. Harrison. 2014. Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. Proc Natl Acad Sci U S A 111:9205-9210.

Casewell, N. R., W. Wuster, F. J. Vonk, R. A. Harrison, and B. G. Fry. 2013. Complex cocktails: the evolutionary novelty of venoms. Trends Ecol Evol 28:219-229.

Castoe, T. A., A. P. de Koning, K. T. Hall, D. C. Card, D. R. Schield, M. K. Fujita, R. P. Ruggiero, J. F. Degner, J. M. Daza, W. Gu, J. Reyes-Velasco, K. J. Shaney, J. M. Castoe, S. E. Fox, A. W. Poole, D. Polanco, J. Dobry, M. W. Vandewege, Q. Li, R. K. Schott, A. Kapusta, P. Minx, C. Feschotte, P. Uetz, D. A. Ray, F. G. Hoffmann, R. Bogden, E. N. Smith, B. S. Chang, F. J. Vonk, N. R. Casewell, C. V. Henkel, M. K. Richardson, S. P. Mackessy, A. M. Bronikowski, M. Yandell, W. C. Warren, S. M. Secor, and D. D. Pollock. 2013. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. Proc Natl Acad Sci U S A 110:20645-20650.

Daltry, J. C., W. Wuster, and R. S. Thorpe. 1996. Diet and snake venom evolution. Nature 379:537-540.

Dowell, N. L., M. W. Giorgianni, V. A. Kassner, J. E. Selegue, E. E. Sanchez, and S. B. Carroll. 2016. The Deep Origin and Recent Loss of Venom Toxin Genes in Rattlesnakes. Curr Biol 26:2434-2445.

423 Fetzner, J. W., Jr. 1999. Extracting high-quality DNA from shed reptile skins: a
424 simplified method. Biotechniques 26:1052-1054.
425 Finn, R. D., P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S.
426 C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G. A. Salazar, J.
427 Tate, and A. Bateman. 2016. The Pfam protein families database:
428 towards a more sustainable future. Nucleic Acids Research 44:6.
429 Fry, B. G. 2005. From genome to "venome": molecular origin and evolution of
430 the snake venom proteome inferred from phylogenetic analysis of toxin
431 sequences and related body proteins. Genome Res 15:403-420.
432 Fry, B. G., H. Scheib, L. van der Weerd, B. Young, J. McNaughtan, S. F.
433 Ramjan, N. Vidal, R. E. Poelmann, and J. A. Norman. 2008. Evolution
434 of an arsenal: structural and functional diversification of the venom
435 system in the advanced snakes (Caenophidia). Mol Cell Proteomics
436 7:215-246.
437 Gelfman S, Burstein, D., Penn, O., Savchenko, A. et al. 2012. Changes in
438 exon-intron structure during vertebrate evolution affect the splicing
439 pattern of exons. Genome Res 22: 35-50.
440 Hargreaves, A. D., M. T. Swain, M. J. Hegarty, D. W. Logan, and J. F. Mulley.
441 2014. Restriction and recruitment-gene duplication and the origin and
442 evolution of snake venom toxins. Genome Biol Evol 6:2088-2095.
443 Heise, P. J., L. R. Maxson, H. G. Dowling, and S. B. Hedges. 1995. Higher-
444 level snake phylogeny inferred from mitochondrial DNA sequences of
445 12S rRNA and 16S rRNA genes. Mol Biol Evol 12:259-265.
446 Hu, B., J. Jin, A. Y. Guo, H. Zhang, J. Luo, and G. Gao. 2015. GSDS 2.0: an
447 upgraded gene feature visualization server. Bioinformatics 31:1296-
448 1297.
449 Jayanthi, G. P. and T. V. Gowda. 1988. Geographical variation in India in the
450 composition and lethal potency of Russell's viper (Vipera russelli)
451 venom. Toxicon 26:257-264.
452 Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass and M. J. E. Sternberg.
453 2015. The Phyre2 web portal for protein modeling, prediction and
454 analysis. Nature Protocols 10: 845–858.
455 Kerkkamp, H. M., R. M. Kini, A. S. Pospelov, F. J. Vonk, C. V. Henkel, and M.
456 K. Richardson. 2016. Snake Genome Sequencing: Results and Future
457 Prospects. Toxins (Basel) 8.
458 Luo, R., B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, Y.
459 Liu, J. Tang, G. Wu, H. Zhang, Y. Shi, Y. Liu, C. Yu, B. Wang, Y. Lu, C.
460 Han, D. W. Cheung, S. M. Yiu, S. Peng, Z. Xiaoqian, G. Liu, X. Liao, Y.
461 Li, H. Yang, J. Wang, T. W. Lam, and J. Wang. 2012. SOAPdenovo2:
462 an empirically improved memory-efficient short-read de novo
463 assembler. Gigascience 1:18.
464 Malhotra, A., S. Creer, C. E. Pook, and R. S. Thorpe. 2010. Inclusion of
465 nuclear intron sequence data helps to identify the Asian sister group of
466 New World pitvipers. Mol Phylogenet Evol 54:172-178.
467 Mallow, D., D. Ludwig, and G. Nilson. 2003. True vipers : natural history and
468 toxinology of Old World vipers. Krieger Pub. Co., Malabar, Fla.
469 Mohapatra, B., D. A. Warrell, W. Suraweera, P. Bhatia, N. Dhingra, R. M.
470 Jotkar, P. S. Rodriguez, K. Mishra, R. Whitaker, P. Jha, and C. Million
471 Death Study. 2011. Snakebite mortality in India: a nationally
472 representative mortality survey. PLoS Negl Trop Dis 5:e1018.

473 Rokyta, D. R., K. P. Wray, A. R. Lemmon, E. M. Lemmon, and S. B. Caudle.
474 2011. A high-throughput venom-gland transcriptome for the Eastern
475 Diamondback Rattlesnake (Crotalus adamanteus) and evidence for
476 pervasive positive selection across toxin classes. Toxicon 57:657-671.
477 Schmieder, R. and R. Edwards. 2011. Fast identification and removal of
478 sequence contamination from genomic and metagenomic datasets.
479 PLoS One 6:e17288.
480 Ullate-Agote, A., M. C. Milinkovitch, and A. C. Tzika. 2014. The genome
481 sequence of the corn snake (Pantherophis guttatus), a valuable
482 resource for EvoDevo studies in squamates. Int J Dev Biol 58:881-888.
483 Vicoso, B., J. J. Emerson, Y. Zektser, S. Mahajan, and D. Bachtrog. 2013.
484 Comparative sex chromosome genomics in snakes: differentiation,
485 evolutionary strata, and lack of global dosage compensation. PLoS Biol
486 11:e1001643.
487 Vonk, F. J., N. R. Casewell, C. V. Henkel, A. M. Heimberg, H. J. Jansen, R. J.
488 McCleary, H. M. Kerkkamp, R. A. Vos, I. Guerreiro, J. J. Calvete, W.
489 Wuster, A. E. Woods, J. M. Logan, R. A. Harrison, T. A. Castoe, A. P.
490 de Koning, D. D. Pollock, M. Yandell, D. Calderon, C. Renjifo, R. B.
491 Currier, D. Salgado, D. Pla, L. Sanz, A. S. Hyder, J. M. Ribeiro, J. W.
492 Arntzen, G. E. van den Thillart, M. Boetzer, W. Pirovano, R. P. Dirks,
493 H. P. Spaink, D. Duboule, E. McGlinn, R. M. Kini, and M. K.
494 Richardson. 2013. The king cobra genome reveals dynamic gene
495 evolution and adaptation in the snake venom system. Proc Natl Acad
496 Sci U S A 110:20651-20656.
497 Vonk, F. J., K. Jackson, R. Doley, F. Madaras, P. J. Mirtschin, and N. Vidal.
498 2011. Snake venom: From fieldwork to the clinic: Recent insights into
499 snake biology, together with new technology allowing high-throughput
500 screening of venom, bring new hope for drug discovery. Bioessays
501 33:269-279.
502 Whitaker, R. 2015. Snakebite in India today. Neurol India 63:300-303.
503 Wüster, W., L. Peppin, C. E. Pook, and D. E. Walker. 2008. A nesting of
504 vipers: Phylogeny and historical biogeography of the Viperidae
505 (Squamata: Serpentes). Mol Phylogenet Evol 49:445-459.
506 Yates, C. M., I. Filippis, L. A. Kelley and M. J. Sternberg. 2014. SuSPect:
507 Enhanced Prediction of Single Amino Acid Variant (SAV) Phenotype
508 Using Network Features J Mol Biol. 426(14): 2692-2701
509 Yin, W., Z. J. Wang, Q. Y. Li, J. M. Lian, et al. 2016. Evolutionary trajectories
510 of snake genes and genomes revealed by comparative analyses of
511 five-pacer viper. Nat Commun 7:13107.

512 **Table 1:** Venom-associated genes and their representative families used in

513 the current study.

514

Gene	Species with the available sequence information	Protein Family
ACHE	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus* and *Thamnophis sirtalis*	Acetylcholinesterase
ADAM11		ADAM (disintegrin/ metalloprotease)
ADAM17		
ADAM19		
ADAM23		
PROK1	*Protobothrops mucrosquamatus, Ophiophagus hannah* and *Python bivitattus*	AVIT (prokinectin)
PROK2	*Protobothrops mucrosquamatus and Ophiophagus hannah, Python bivittatus* and *Thamnophis sirtalis*	
CPAMD8	*Protobothrops mucrosquamatus, Python bivittatus* and *Thamnophis sirtalis*	Complement C3
crotasin	*Protobothrops mucrosquamatus*	Crotasin/ beta defensin
CST1	No sequence information is available in any of the four species	Cystatin
CST3	*Ophiophagus hannah*	
CST4	No sequence information is available in any of the four species	
CSTA	*Protobothrops mucrosquamatus* and *Thamnophis sirtalis*	
EDN1	*Protobothrops mucrosquamatus* and *Python bivittatus*	Endothelin
EDN3	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus*, *Thamnophis sirtalis* and	
F5		Factor V
F10	*Ophiophagus hannah*	Factor X
KLKB1	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus, Thamnophis sirtalis* and	Kallikrein
KLK14	*Ophiophagus hannah*	
kunitoxin	*Protobothrops mucrosquamatus, Python bivittatus* and *Ophiophagus hannah*	Kunitz-type protease inhibitor
LYNX1	*Ophiophagus hannah*	LYNX/SLUR
CLEC3A	*Protobothrops mucrosquamatus, Python bivittatus* and *Thamnophis sirtalis*	Lectin
CLEC3B	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus* and *Thamnophis sirtalis*	
CLEC11A	*Protobothrops mucrosquamatus, Python bivittatus* and *Thamnophis sirtalis*	
CLEC16A	*Protobothrops mucrosquamatus, Python bivittatus* and *Thamnophis sirtalis*	
CLEC19A	*Protobothrops mucrosquamatus and Python bivittatus*	
NPR1	*Protobothrops mucrosquamatus, Python bivittatus* and *Thamnophis sirtalis*	Natriuretic peptide
NPR2	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus* and *Thamnophis sirtalis*	
NPR3		
NGF	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus, Thamnophis sirtalis, Protobothrops flavoviridis, Crotalus horridus, Sistrurus miliarius barbouri* and *Boa constrictor*	Beta-nerve growth factor
PLAA	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus* and *Thamnophis sirtalis*	Phospholipase A (2)
PLA2R1		

PLA2G1B	*Python bivittatus, Thamnophis sirtalis* and *Protobothrops mucrosquamatus*	
PLA2G10	*Protobothrops mucrosquamatus, Protobothrops flavoviridis, Thamnophis sirtalis, Ophiophagus hanna* and *Python bivittatus*	
PLA2G12A	*Python bivittatus, Thamnophis sirtalis* and *Protobothrops mucrosquamatus*	
PLA2G12B	*Protobothrops mucrosquamatus, Ophiophagus hannah* and *Python bivittatus*	
PLA2G15	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus and Thamnophis sirtalis*	
PLA2G3	*Protobothrops mucrosquamatus, Python bivittatus* and *Ophiophagus hannah*	
PLA2G4A	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus* and *Thamnophis sirtalis*	
PLA2G4C	*Protobothrops mucrosquamatus, Python bivittatus* and *Thamnophis sirtalis*	
PLA2G6	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus* and *Thamnophis sirtalis*	
PLA2G7		
SPSB4	*Protobothrops mucrosquamatus* and *Thamnophis sirtalis*	SPIa/Ryanodine
SPSB3	*Protobothrops mucrosquamatus, Python bivittatus* and *Thamnophis sirtalis*	
SPSB1	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus* and *Thamnophis sirtalis*	
VEGFA1	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus, Thamnophis sirtalis, Crotalus horridus* and *Protobothrops flavoviridis*	Vascular endothelial growth factor (VEGF)
VEGFA2		
VEGFA3		
VEGFB	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus, Thamnophis sirtalis, Crotalus horridus, Protobothrops flavoviridis* and *Sistrurus miliarius barbouri*	
VEGFC	*Protobothrops mucrosquamatus, Python bivittatus* and *Thamnophis sirtalis*	
VEGFF	*Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus, Thamnophis sirtalis* and *Protobothrops flavoviridis*	
WAP	*Protobothrops mucrosquamatus, Python bivittatus, Thamnophis sirtalis* and *Ophiophagus hannah*	Whey acidic protein/secretory leukoproteinase inhibitor
WFIKKN1		
WFIKKN2		
CRISP	*Protobothrops flavoviridis, Protobothrops mucrosquamatus, Ophiophagus hannah, Python bivittatus, Thamnophis sirtalis, Crotalus horridus, Calloselasma rhodostoma, Sistrurus miliarius barbouri* and *Deinagkistrodon acutus*	CRISP

515

516

517 **Figure Legends**

518 **Fig. 1**: Comparative analyses of nerve growth factor (NGF). NGF gene,
519 mRNA, and protein domains of Russell's viper (A) and its comparison with the
520 consensus NGF sequences from all five vertebrate phyla (fishes, amphibians,
521 reptiles, aves and mammals) (B), with venomous (V) vertebrates from multiple
522 phyla of vertebrates and invertebrates (C), and from various reptilian
523 subgroups (D). The shades of brown and grey in B and C represent
524 conservation to various degrees and variability, respectively. Grey in D
525 represents conserved residues, red represents variable residues in New
526 World vipers (NWV), yellow and green represent conserved and variable
527 residues in Old World vipers (OWV), and elapids respectively.

528

529 **Fig. 2:** Comparative analyses of vascular endothelial growth factor - A (VEGF-
530 A). Organization of the gene, mRNA, and protein domains of Russell's viper
531 PDGF domain (A) and its comparison with the consensus sequences from all
532 five vertebrate phyla (fishes, amphibians, reptiles, aves and mammals) (B),
533 from the venomous (V) vertebrates and invertebrates (C), and from various
534 reptilian subgroups (D). The shades of brown and grey in B and C represent
535 conserved and varying residues, respectively. Grey in D represents conserved
536 residues, red represents variable residues in New World vipers (NWV), yellow
537 and green represent conserved and variable residues in Old World vipers
538 (OWV), and elapids respectively

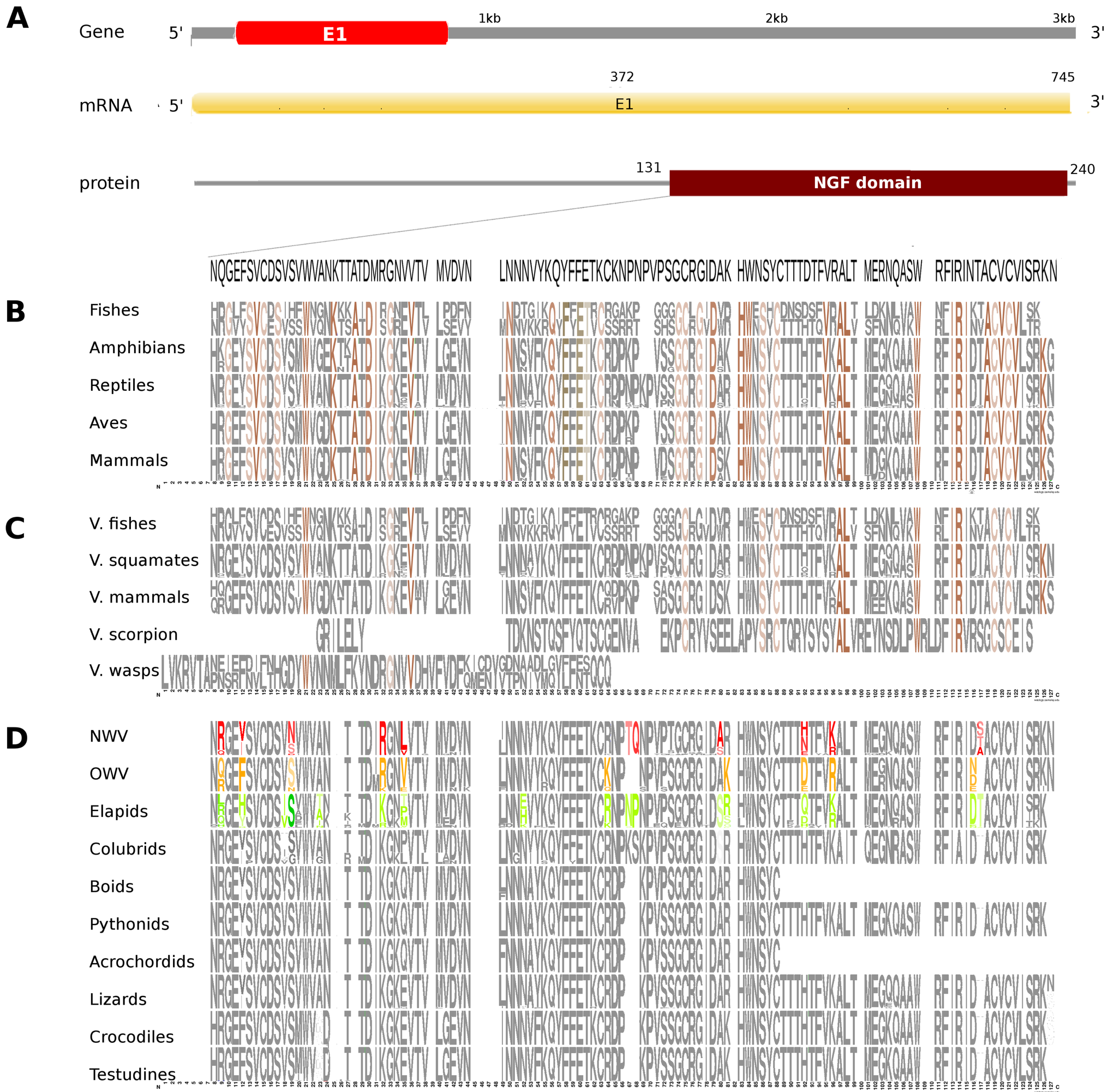
539

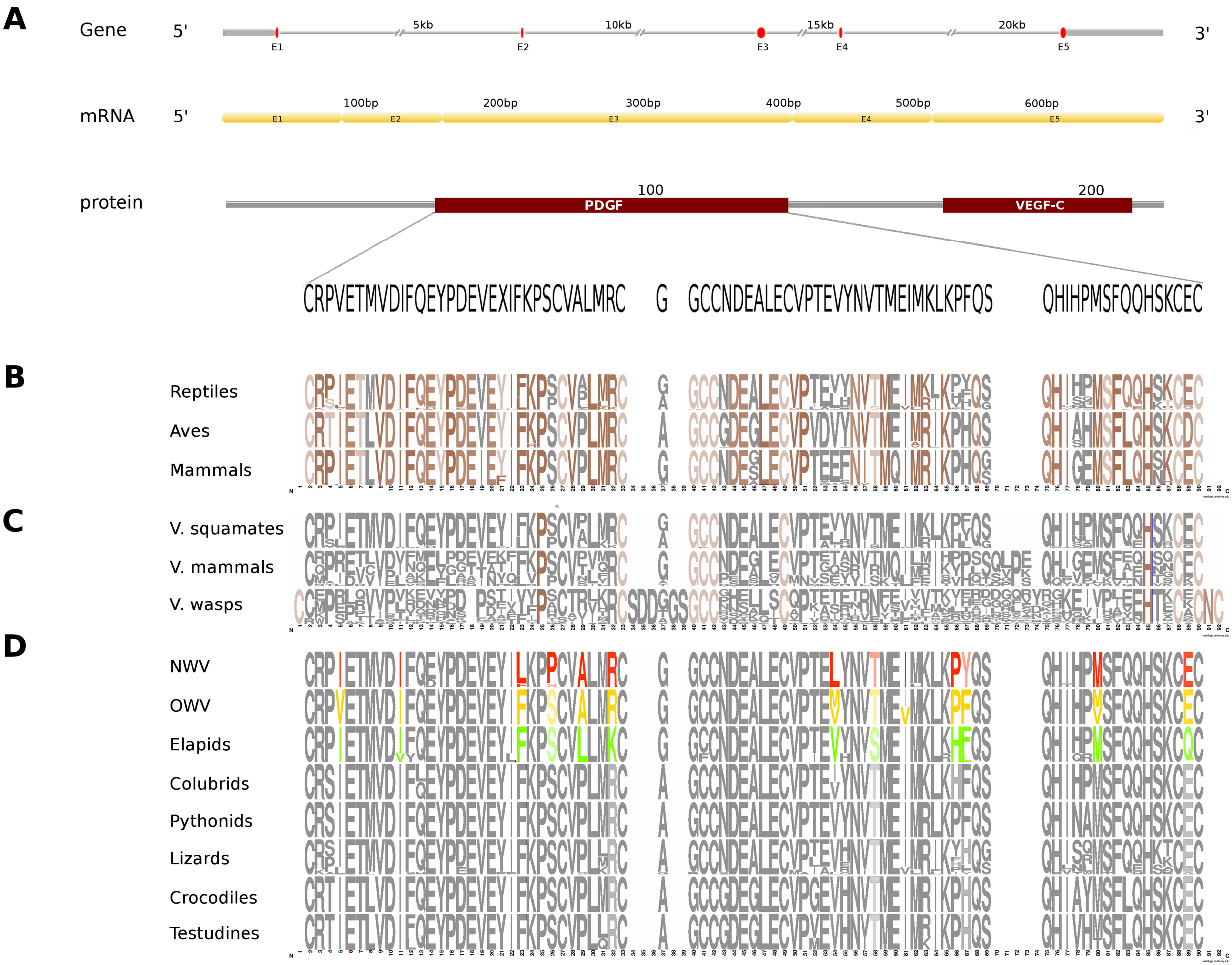
540 **Fig. 3:** Comparative analyses of kunitoxin. Organization of the gene, mRNA,
541 and protein domains of Russell's viper (A) and its comparison with the

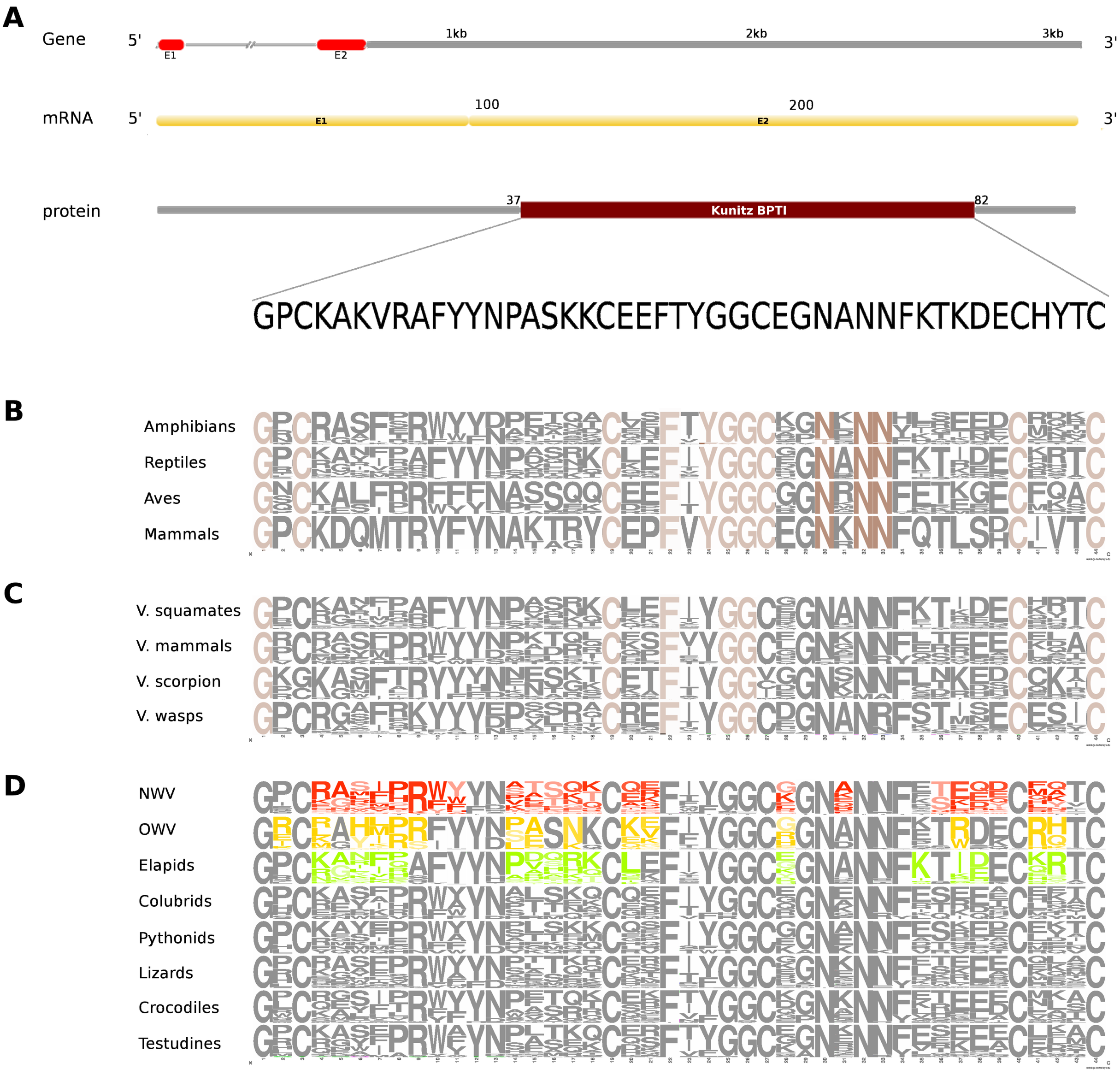
542 consensus BPTI domain sequences from all five vertebrate phyla (fishes,
543 amphibians, reptiles, aves and mammals) (B), from venomous (V) vertebrates
544 and invertebrates (C), from various reptilian subgroups (D). The shades of
545 brown and grey in B and C represent conserved and varying residues,
546 respectively. Grey in D represents conserved residues, red represents
547 variable residues in New World vipers (NWV), yellow and green represent
548 conserved and variable residues in Old World vipers (OWV), and elapids
549 respectively.
550
551 **Fig. 4:** Comparative analyses of CRISP. Organization of CRISP gene, mRNA,
552 and protein domains of Russell's viper (A) and its comparison with the
553 consensus CRISP sequences from all five vertebrate phyla (fishes,
554 amphibians, reptiles, aves and mammals, B and E); from venomous animals
555 (V) vertebrates (fishes, squamates and mammals) and invertebrates
556 (scorpions and wasps, C and F); and from various reptilian subgroups (D and
557 G). The shades of brown and grey in B, C, E and F represent conserved and
558 varying residues, respectively. Grey in D and G represents conserved
559 residues, red represents variable residues in New World vipers (NWV), yellow
560 and green conserved and variable residues in Old World vipers (OWV), and
561 elapids respectively.
562
563 **Fig. 5:** Three-dimensional protein structural models to access quality and
564 functional differences in key venom-associated proteins (NGF, A; PDGF, B;
565 Kunitz BPTI, C; CAP, D; and CRISP, E) across New World vipers (NWV), Old

566 World vipers (OWV) and elapids. The status of the parameters being
567 investigated using Phyre2 are indicated in the color legends on the side.
568
569 **Supplementary File legends**
570 **Table S1:** Genbank accession numbers for Russell's viper venom-associated
571 proteins.
572
573 **Table S2:** The amino acid sequences for the Russell's viper's toxin-
574 associated genes.
575
576 **Table S3:** List of completely sequenced snake species.
577
578 **Table S4:** Accession IDs and taxonomic groups of species used in all
579 comparative analyses.
580
581 **File S1:** Agarose gel electrophoresis of genomic DNA extracted from skin
582 exuviate of Russell's viper.
583 .
584 **File S2:** Exon-intron structures for all the toxin-associated genes in Russell's
585 viper.
586
587 **File S3:** Graphical summary of comparative domain analysis performed for all
588 toxin-associated genes across 18 completely sequenced snake species.
589

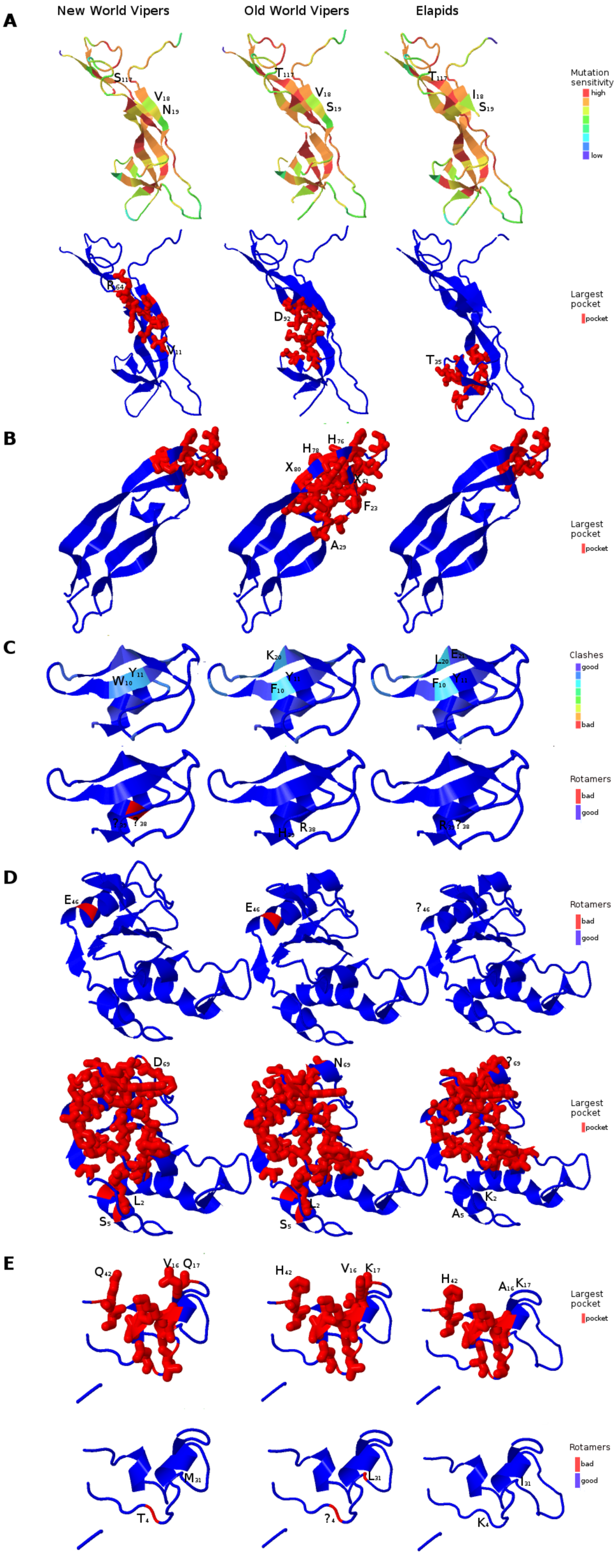
590 **File S4:** A larger pool of sequences available from various snakes and other
591 reptile groups (Colubridae, Boidae, Pythonidae, Acrochordidae, Lizards,
592 Crocodiles and Testudines) were compared with the sequence of Russell's
593 viper for NGF, PDGF, Kunitz BPTI, CAP and CRISP domains.
594
595 **File S5:** 3D protein structural model quality and function measuring parameter
596 comparison across New World vipers (NWV), Old World vipers (OWV) and
597 elapids in NGF, PDGF, Kunitz BPTI, CAP and CRISP domains. The status of
598 the parameters being investigated using Phyre2 are indicated in the color
599 legends on the side.


A
Gene
5'
E1
1kb
2kb
3kb
3'
mRNA
5'
372
745
E1
3'
protein
131
240
NGF domain
B
Fishes
Amphibians
Reptiles
Aves
Mammals
C
V. fishes
V. squamates
V. mammals
V. scorpion
V. wasps
D
NWV
OWV
Elapids
Colubrids
Boids
Pythonids
Acrochordids
Lizards
Crocodiles
Testudines


A
Gene
5'
3'
5kb
10kb
15kb
20kb
E1
E2
E3
E4
E5
mRNA
100bp
200bp
300bp
400bp
500bp
600bp
protein
100
200
PDGF
VEGF-C
CRPVETMVDIFQEYPDEVEXIFKPSCVALMRC G GCCNDEALECVPTEVYNVTMEIMKLKPFQS QHIHPMSFQQHSKCEC
B
Reptiles
Aves
Mammals
C
V. squamates
V. mammals
V. wasps
D
NWV
OWV
Elapids
Colubrids
Pythonids
Lizards
Crocodiles
Testudines


A
Gene
5'
E1
E2
1kb
2kb
3kb
3'
mRNA
5'
100
200
E1
E2
3'
protein
37
Kunitz BPTI
82
GPCKAKVRAFYYNPASKKCEEFTYGGCEGNANNFKTKDECHYTC
B
Amphibians
Reptiles
Aves
Mammals
C
V. squamates
V. mammals
V. scorpion
V. wasps
D
NWV
OWV
Elapids
Colubrids
Pythonids
Lizards
Crocodiles
Testudines

A
Gene
5'
3'
5kb
10kb
15kb
20kb
25kb
E1
E2
E3
E4
E5
E6
E7
E8
mRNA
100
200
300
400
500
600
700
Protein
CAP
CRISP
B
Fishes
Amphibians
Reptiles
Aves
Mammals
C
V. Amphibians
V. squamates
V. mammals
V. scorpion
V. wasps
D
NWV
OWV
Elapids
Colubrids
Pythonids
Lizards
Crocodiles
Testudines
E
F
V. squamates
V. mammals
G


A
New World Vipers
Old World Vipers
Elapids
S117
V18
N19
T117
V18
S19
T117
I18
S19
Mutation sensitivity
high
low
R64
V11
D52
T35
Largest pocket
pocket
B
H78
H76
X80
X61
F23
A29
Largest pocket
pocket
C
W10
Y11
K20
Y11
F10
L20
E21
F10
Y11
Clashes
good
bad
Rotamers
bad
good
D
E46
E46
?46
Rotamers
bad
good
D69
N69
?69
L2
S5
L2
S5
K2
A5
Largest pocket
pocket
E
Q42
V16
Q17
H42
V16
K17
H42
A16
K17
Largest pocket
pocket
M31
T4
L31
?4
I31
K4
Rotamers
bad
good